

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

Identificação e avaliação do impacto de
outliers em processos
ARFIMA (p, d, q)

AUTOR: MATHIAS MARTENS
ORIENTADOR: PROFESSOR DR. CLEBER BISOGNIN

Porto Alegre, dezembro de 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

DEPARTAMENTO DE ESTATÍSTICA

Identificação e avaliação do impacto de

outliers em processos

$\text{ARFIMA}(p, d, q)$

MATHIAS MARTENS

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:

Professor Dr. Cleber Bisognin

Professor Dr. Márcio Valk

Porto Alegre, dezembro de 2014.

Agradecimentos

Primeiramente agradeço à minha família, em especial aos meus pais Selmo e Heidi, aos meus irmãos Marcio, Mauro e Mara, às cunhadas Cristina e Danusa, por me apoiarem em todas decisões ao longo desse período, por sempre me estenderem a mão nos momentos em que precisei e principalmente pelo carinho.

Aos mestres que não pouparam esforços em compartilhar seus conhecimentos e experiências, colaborando com o meu crescimento intelectual, em especial ao meu orientador Cleber Bisognin e a professora Lisiane Selau pelo apoio, amizade e conhecimento compartilhado durante o curso.

Aos colegas de curso, que compartilharam todos os momentos de alegria e angústia enfrentados, em especial aos colegas Gustavo Thomas e Tiago Jantsch, amigos e companheiros desde o início.

À todos meus amigos que sempre me apoiaram, mesmo estando longe, em especial aos irmãos de coração Tiago Lorenzoni e Gustavo Heineck por me acompanharem desde a infância.

À minha namorada Mariana Pies Gionbelli pelo amor, carinho e companheirismo em todos os momentos.

Aos colegas e amigos das empresas SPHINX Brasil e Statsoft, por todo conhecimento e crescimento profissional que me proporcionaram durante os períodos de estágio e por sempre acreditarem no meu potencial, em especial a Henrique Freitas, Fernando Kuhn Andriotti, Ricardo Simm Costa, Pedro Oliveira, Katrine Guewehr e Taiane Prass.

Por último, mas não menos importante, agradeço à UFRGS por proporcionar a oportunidade de ter um estudo de qualidade e ainda morar na CEU, possibilitando a minha permanência em Porto Alegre.

MUITO OBRIGADO.

Resumo

Atualmente um dos problemas mais importantes no estudo de séries temporais é a presença de *outliers*, isto é, quando observações são afetadas por erros grosseiros ou fatores externos, como por exemplo, enchentes, greves e mudanças políticas ou econômicas. Este problema afeta negativamente a estimação dos parâmetros do modelo e, conseqüentemente, a previsão de observações futuras. Neste trabalho, apresentamos os estimadores paramétricos por verossimilhança aproximada, considerando a série livre de *outliers* e quando contaminada por *outliers* aditivos e inovadores. Ainda, estendemos a metodologia de detecção de *outliers* em modelos ARMA (p, q) , denotada por *SODA*, introduzido por Choy (2001) para modelos ARFIMA (p, d, q) , caracterizados por longa dependência. Comparamos esse procedimento com o método do *Teste da Razão de Verossimilhança* proposto por Bisognin (2007) para modelos SARFIMA $(p, d, q) \times (P, D, Q)_S$. Para testar e avaliar os respectivos métodos foram utilizadas simulações de Monte Carlo, onde conhecemos a priori os parâmetros do modelo, a magnitude de perturbação e a posição ou a probabilidade de contaminação dos *outliers*, para então verificar a eficiência dos métodos e o impacto que a correção nos *outliers* tem nas estimativas dos parâmetros do modelo. Como resultado obtemos um indicativo de maior eficiência na estimação da magnitude de perturbação dos *outliers* através do método do *Teste da Razão de Verossimilhança*, tanto para *outliers* aditivos quanto inovadores, apesar de o método *SODA* também apresentar bons resultados. Concluimos então que os métodos propostos são eficientes na estimação, identificação e correção de *outliers* em processos ARFIMA (p, d, q) .

Abstract

Nowadays one of the most important problems in the study of time series analysis is the presence of outliers, observations changed by input errors or external factors as floods, strikes and political or economic changes. This problem affect the estimation of the model parameters and consequently, the prediction of future observations negatively. In this paper we present parametric estimators by approximate likelihood, considering free outlier series and when contaminated by additive outliers and innovations outliers. Moreover, we extend the outliers detection methodology in ARMA (p, q) models, denoted by *SODA*, introduced by Choy (2001) for ARFIMA (p, d, q) models characterized by long memory. We compare this procedure with the likelihood ratio test, suggested by Bisognin (2007) for SARFIMA $(p, d, q) \times (P, D, Q)_S$ models. To test and evaluate those methods, Monte Carlo simulations were used since we already know the model parameters, the perturbation magnitude and the position or the probability of outlier contamination. Then, we verify the efficiency of the methods and the impact that outlier correction has in the model parameters estimate. As result we obtain signals of greater efficiency in the perturbation magnitude of outliers over likelihood ratio test method for additive outliers and innovations outliers, although the *SODA* method also present good results. We conclude that the proposed methods are efficient in the estimation, identification and correction of outliers in ARFIMA (p, d, q) process.

ÍNDICE

1	Introdução	9
2	Processos ARFIMA(p, d, q)	11
2.1	Definições básicas	11
2.2	Propriedades do processo ARFIMA (p, d, q)	12
3	Métodos de contaminação	16
3.1	Paramétrico	16
3.2	Mistura	18
4	Estimação e Identificação de <i>outliers</i>	20
4.1	Estimação	21
4.1.1	<i>FT</i> sem contaminação	21
4.1.2	<i>FT</i> contaminado por <i>outliers</i> aditivos	22
4.1.3	<i>FT</i> contaminado por <i>outliers</i> inovadores	22
4.2	Identificação de <i>outliers</i>	23
4.2.1	<i>Tsoutlier</i>	23
4.2.2	<i>SODA</i>	24
4.2.3	<i>Teste da razão de verossimilhança</i>	26
5	Resultados por simulação	30
5.1	Resultados do método <i>tsoutlier</i>	30
5.2	Resultados de Estimação pelos métodos <i>FT</i> , <i>FTAO</i> e <i>FTIO</i> .	41
5.3	Resultados para os métodos <i>SODA</i> e <i>Teste da Razão de Verossimilhança</i>	48
5.4	Exemplo de procedimento completo	56
6	Conclusão	59
	Referências	61
	Apêndice A	65

LISTA DE FIGURAS

5.1	Processos ARFIMA (p, d, q) contaminados pelo método paramétrico nas posições $(50, 100, 150, 200)$, onde $p = 0$, $q = 0$, $d = 0.3$, $n = 500$ e $w \in \{4, 7, 10\}$	31
5.2	Processo ARFIMA (p, d, q) contaminados pelo método paramétrico nas posições $(50, 100, 200, 500)$, onde $p = 0 = q$, $d = 0.3$, $n = 1000$ e $w = 10$	56

LISTA DE TABELAS

5.1	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.1$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	32
5.2	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	33
5.3	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.45$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	34
5.4	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.1$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	35
5.5	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.3$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	36
5.6	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.45$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	37
5.7	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.1$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	38
5.8	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.3$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	39

5.9	Número de séries que o método <i>tsoutliers</i> classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por <i>AO</i> e <i>IO</i> , quando $d = 0.45$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	40
5.10	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	42
5.11	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 1$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	43
5.12	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	44
5.13	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	45
5.14	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 1$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	46
5.15	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	47
5.16	Estimação da magnitude da perturbação w , quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$	50
5.17	Estimação da magnitude da perturbação w , quando $d = 0.3$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$	51
5.18	Estimação da magnitude da perturbação w , quando $d = 0.3$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$	52
5.19	Estatísticas descritivas das estatísticas de detecção de <i>outliers</i> quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$	54
5.20	Estatísticas descritivas das estatísticas de detecção de <i>outliers</i> quando $d = 0.3$, $p = 1$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$	55
5.21	Estatísticas descritivas das estatísticas de detecção de <i>outliers</i> quando $d = 0.3$, $p = 0$, $q = 1$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$	55
5.22	Estimativas do parâmetro d para o processo ARFIMA (p, d, q) contaminado por <i>AO</i> quando $p = 0 = q$ e $d = 0.3$	56
5.23	Estimativas da magnitude de perturbação w e sua estatística de detecção Γ para o processo ARFIMA (p, d, q) contaminado por <i>AO</i> quando $p = 0 = q$ e $\hat{d} = 0.3287$	57
5.24	Estimativas do parâmetro d para o processo ARFIMA (p, d, q) contaminado por <i>AO</i> quando $p = 0 = q$ e $d = 0.3$ após aplicar o método <i>SODA</i>	57
5.25	Estimativas da magnitude de perturbação w e sua estatística de detecção λ para o processo ARFIMA (p, d, q) contaminado por <i>AO</i> quando $p = 0 = q$ e $\hat{d} = 0.3287$	58
5.26	Estimativas do parâmetro d para o processo ARFIMA (p, d, q) contaminado por <i>AO</i> quando $p = 0 = q$ e $d = 0.3$ após aplicar o método <i>Teste da Razão de Verossimilhança</i>	58

A.1	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1, p = 0, q = 0, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	66
A.2	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45, p = 0, q = 0, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	67
A.3	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1, p = 1, \phi_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	68
A.4	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45, p = 1, \phi_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	69
A.5	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1, p = 1, \theta_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	70
A.6	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45, p = 1, \theta_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	71
A.7	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1, p = 0, q = 0, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$	72
A.8	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45, p = 0, q = 0, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	73
A.9	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1, p = 1, \phi_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	74
A.10	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45, p = 1, \phi_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	75
A.11	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1, p = 1, \theta_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	76
A.12	Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45, p = 1, \theta_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$. . .	77

Capítulo 1

Introdução

Estudos de séries temporais são amplamente utilizados em diversas áreas, onde o objetivo é analisar o comportamento de uma série de observações ao longo do tempo e, principalmente, fazer previsões para o futuro. Existem duas abordagens para tal análise, domínio do tempo e domínio da frequência, onde no domínio do tempo é analisado principalmente as propriedades de autocovariância e autocorrelação, enquanto no domínio da frequência são utilizados métodos com base na função densidade espectral.

Além de duas abordagens para a análise, temos mais duas definições muito importantes para modelagem, que são as propriedades de curta e longa dependência. Para cada uma das propriedades descritas foram propostos diferentes modelos, sendo um dos mais famosos o modelo de curta dependência *auto-regressivo integrado de média móvel*, denotado por $ARIMA(p, d, q)$, proposto inicialmente por Box e Jenkins na década de 70.

Granger e Joyeux (1980) e Hosking (1981 e 1984) definem o modelo *auto-regressivo fracionariamente integrado de média móvel*, denotado por $ARFIMA(p, d, q)$ que é um modelo que possui a propriedade de longa dependência para séries estacionárias, obtido através de uma extensão natural do modelo $ARMA(p, q)$.

Um problema recorrente em dados de séries temporais é a presença de erros grosseiros e influências externas, como por exemplo, enchentes, greves e mudanças políticas ou econômicas. Estes fatos, comumente chamados de outliers, representam uma preocupação quanto a eficiência das estimativas dos parâmetros dos modelos e, conseqüentemente, a previsão de observações futuras. Diversos autores como Chen e Liu (1993), Fox (1972), Tsay (1986), Abraham e Box (1979), Choy (2001), abordam este problema.

Fox (1972) introduziu os conceitos de outliers do tipo I e tipo II, conhecidos na literatura como outliers aditivos e inovadores, denotados por AO e IO , respectivamente. Outliers aditivos são observações discrepantes que afetam um único momento no tempo, por exemplo, um erro de digitação. Por outro lado, outliers inovadores afetam não somente uma observação, mas uma série de observações em um certo intervalo de tempo, por exemplo, uma

enchente em um rio.

Ainda, surgem questões referentes aos modelos geradores desses *outliers*, como o modelo de contaminação por mistura descrito por Denby e Martin (1979) e Bustos e Yohai (1986) e o modelo de contaminação paramétrico descrito por Hotta e Neves (1992) e Palma (1998).

Quando existem *outliers* na série temporal, se for utilizado o estimador de máxima verossimilhança aproximada, denotado por FT , descrito por Fox e Taqqu (1986) e Whittle (1951), já sabemos que as estimativas dos parâmetros é afetada negativamente (veja por exemplo, Martin e Yohai 1986 e Peña 1990). Considerando o modelo de contaminação por mistura para processos SARFIMA $(p, d, q) \times (P, D, Q)_S$, Bisognin (2007) propõe uma alteração para o estimador FT nos casos onde a série está contaminada por *outliers* aditivos e inovadores. Esta alteração consiste em introduzir mais um parâmetro para ser estimado pelo método de máxima verossimilhança, resultando em uma estimação mais precisa dos parâmetros do modelo.

Este trabalho tem como objetivo estender de modelos ARMA para modelos ARFIMA o método *SODA*, aperfeiçoar e comparar os métodos *SODA* e *Teste da Razão de Verossimilhança* para detectar a posição, o tipo e a magnitude de *outliers* dos tipos *AO* e *IO*, através do modelo de contaminação paramétrico e do modelo de contaminação por mistura.

No Capítulo 2, apresentamos propriedades importantes dos processos ARFIMA (p, d, q) . No Capítulo 3 apresentamos os métodos de contaminação onde a Seção 3.1 apresenta o modelo de contaminação paramétrico e a Seção 3.2 apresenta o modelo de contaminação por mistura. No Capítulo 4, apresentamos os métodos de estimação por máxima verossimilhança aproximada FT , $FTAO$ e $FTIO$ na Seção 4.1, seguidos da Seção 4.2 onde apresentamos os métodos *tsoutlier*, *SODA* e o *Teste da Razão de Verossimilhança*. No Capítulo 5, apresentamos resultados por simulação de Monte Carlo para os três métodos descritos no Capítulo 4 e, por fim, no Capítulo 6 apresentamos as conclusões.

Capítulo 2

Processos ARFIMA(p, d, q)

Séries com propriedade de longa dependência foram encontradas inicialmente pelo hidrólogo Harold E. Hurst em 1951 enquanto investigava a série dos níveis do rio Nilo. Estas séries são caracterizadas pela presença de correlação entre observações separadas por um longo período de tempo.

Inicialmente introduzidos por Granger e Joyeux (1980) e Hosking (1981 e 1984), o modelo *auto-regressivo fracionariamente integrado de média móvel*, denotado por ARFIMA(p, d, q) é uma extensão natural do modelo ARMA(p, q).

2.1 Definições básicas

Nesta seção definimos os processos *auto-regressivos fracionariamente integrados de médias móveis* denotados por ARFIMA(p, d, q) e apresentamos algumas de suas propriedades.

Primeiramente introduzimos algumas definições muito importantes.

Definição 2.1. O processo $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um *ruído branco* com média zero e variância constante, denotado por $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, se

$$\mathbb{E}(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2 \quad e \quad \gamma_\varepsilon(k) = \begin{cases} \sigma_\varepsilon^2 & \text{se } k = 0, \\ 0 & \text{se } k \neq 0. \end{cases}$$

Definição 2.2. Para todo real $d > -1$ definimos o *operador diferença* ∇^d pela expansão binomial

$$\nabla^d \equiv (1 - \mathcal{B})^d = \sum_{k \geq 0} \binom{d}{k} (-\mathcal{B})^k = 1 - d\mathcal{B} - \frac{d(1-d)}{2!} \mathcal{B}^2 - \dots, \quad (2.1)$$

onde

$$\binom{d}{k} \equiv \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)},$$

na qual $\Gamma(\cdot)$ é a função Gama.

A seguir definimos os modelos ARFIMA.

Definição 2.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico que satisfaz a equação

$$\phi(B)\nabla^d(X_t - \mu) = \theta(B)\varepsilon_t, \quad t \in \mathbb{Z} \quad (2.2)$$

onde μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco, ∇^d é o operador diferença (ver Definição 2.2), B é o operador defasagem ou de retardo, isto é, $B^k X_t = X_{t-k}$, d é o *parâmetro ou grau de diferenciação*, $\phi(\cdot)$ e $\theta(\cdot)$ são polinômios de ordem p e q , respectivamente dados por

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \dots - \phi_p z^p = \sum_{i=0}^p \phi_i z^i \\ \theta(z) &= 1 - \theta_1 z - \theta_2 z^2 - \theta_3 z^3 - \dots - \theta_q z^q = \sum_{j=0}^q \theta_j z^j \end{aligned} \quad (2.3)$$

com ϕ_i , $1 \leq i \leq p$, e θ_j , $1 \leq j \leq q$ constantes pertencendo aos reais. Então o processo $\{X_t\}_{t \in \mathbb{R}}$ é chamado de *processo auto-regressivo fracionariamente integrado de média móvel de ordem* (p, d, q) , denotado por ARFIMA (p, d, q) , onde d é o grau ou parâmetro de diferenciação.

O principal motivo para a escolha do modelo ARFIMA (p, d, q) em modelagem de séries temporais com característica de memória longa, é que o efeito do parâmetro d em observações separadas por um longo período de tempo decai hiperbolicamente conforme a distância entre as observações aumenta, portanto d está relacionado com a estrutura de correlação de ordens altas da série. Por outro lado, os efeitos dos parâmetros ϕ e θ decaem exponencialmente, portanto estão relacionados com a estrutura de correlação de ordens baixas.

A seguir vamos apresentar propriedades fundamentais dos modelos ARFIMA $(0, d, 0)$ e ARFIMA (p, d, q) . As demonstrações dos teoremas a seguir podem ser encontradas em Hosking (1981) e Brockwell e Davis (1991).

Para facilitar, utilizaremos $\mu \equiv 0$ sem perda de generalização.

2.2 Propriedades do processo ARFIMA (p, d, q)

Nesta seção, apresentamos as principais propriedades dos processos ARFIMA (p, d, q) .

Teorema 2.1. *Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ruído branco fracionário denotado por ARFIMA $(0, d, 0)$, dado pela Definição 2.3. Então seguem as seguintes afirmações.*

- (i) Quando $d < 0.5$, $\{X_t\}_{t \in \mathbb{Z}}$ é um processo estacionário e tem representação na forma de um processo média móvel infinito dado por

$$X_t = \psi(B)\varepsilon_t = \sum_{k \geq 0} \psi_k \varepsilon_{t-k}, \quad t \in \mathbb{Z}$$

com pesos ψ_k dados por

$$\psi_k = \frac{d(1+d) \dots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!} = \frac{\Gamma(k+D)}{\Gamma(d)\Gamma(k+1)}.$$

Ainda,

$$\psi_k \sim \frac{k^{d+1}}{(d-1)!}, \quad \text{quando } k \rightarrow \infty;$$

- (ii) Quando $d > -0.5$, $\{X_t\}_{t \in \mathbb{Z}}$ é um processo invertível e tem representação na forma de um processo auto-regressivo infinito dado por

$$\sum_{k \geq 0} \pi_k X_{t-k} = \pi(B)X_t = \varepsilon_t, \quad t \in \mathbb{Z}$$

com pesos π_k dados por

$$\pi_k = \frac{-d(1-d) \dots (k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}.$$

Além disso,

$$\pi_k \sim \frac{k^{-d-1}}{(-d-1)!}, \quad \text{quando } k \rightarrow \infty;$$

- (iii) Assumindo que $-0.5 < d < 0.5$, o processo $\{X_t\}_{t \in \mathbb{Z}}$ possui função densidade espectral dada por

$$f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \left(2 \sin\left(\frac{\omega}{2}\right) \right), \quad 0 < \omega \leq \pi. \quad (2.4)$$

Quando $\omega \rightarrow 0$, $f_X(\omega) \sim \omega^{-2d}$.

- (iv) O processo $\{X_t\}_{t \in \mathbb{Z}}$ possui funções de autocovariância e autocorrelação de ordem $k \in \mathbb{N}$, dadas por

$$\gamma_X(k) = \mathbb{E}(X_t, X_{t-k}) = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)} \sigma_\varepsilon^2 \quad (2.5)$$

e

$$\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}. \quad (2.6)$$

Em particular, temos que

$$\rho_X(k) \sim \frac{(-d)!k^{2d-1}}{(d-1)!}, \quad \text{quando } k \rightarrow \infty;$$

(v) O processo $\{X_t\}_{t \in \mathbb{Z}}$ possui função de autocorrelação parcial, dada por

$$\phi_X(k, k) = \frac{d}{k-d}, \quad k \in \mathbb{N}$$

e

$$\phi_X(k, j) = - \binom{k}{j} \frac{\Gamma(j-d)\Gamma(k-d-j+1)}{\Gamma(-d)\Gamma(k-d+1)}, \quad j = 1, \dots, k.$$

O teorema a seguir apresenta as principais propriedades dos processos ARFIMA (p, d, q)

Teorema 2.2. *Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo denotado por ARFIMA (p, d, q) dado pela Definição 2.3, onde*

- $d \in (-0.5, 0.5)$;
- as raízes dos polinômios $\phi(\cdot)$ e $\theta(\cdot)$ estão fora do círculo unitário e não são comuns.

Então,

(i) $\{X_t\}_{t \in \mathbb{Z}}$ é um processo estacionário e tem representação na forma de um processo média móvel infinito dado por

$$X_t = \sum_{k \geq 0} \psi_k \varepsilon_{t-k}$$

onde ψ_k é dado por

$$\psi_k = \frac{\theta(B)}{\phi(B)} (1-B)^{-d}$$

(ii) $\{X_t\}_{t \in \mathbb{Z}}$ é um processo invertível e tem representação na forma de um processo auto-regressivo infinito dado por

$$\sum_{k \geq 0} \pi_k X_{t-k} = \varepsilon_t,$$

onde π_k é dado por

$$\pi_k = \frac{\phi(B)}{\theta(B)} (1 - B)^{-d};$$

(iii) Se $\{X_t\}_{t \in \mathbb{Z}}$ é um processo causal e inversível, sua função densidade espectral é dada por

$$f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \left[2 \sin\left(\frac{\omega}{2}\right)\right]^{-2d}, \quad (2.7)$$

para todo $0 < \omega < \pi$.

(iv) Se $\{X_t\}_{t \in \mathbb{Z}}$ é um processo causal e inversível, sua função de autocorrelação de ordem k é dada por

$$\rho_X(k) \sim U k^{2d-1}, \quad \text{quando } k \rightarrow \infty \quad (2.8)$$

com $U \neq 0$.

Capítulo 3

Métodos de contaminação

Dados de séries temporais, encontrados nas mais diversas áreas, são muitas vezes propensos a fenômenos extremos como, por exemplo, enchentes, greves, mudanças políticas ou econômicas, ondas de calor ou frio inesperadas ou, ainda, erros grosseiros de digitação. Percebemos que essas observações fogem do padrão das demais e são frequentemente nomeadas como *outliers*.

Fox (1972) introduziu os conceitos de *outliers* do tipo I e tipo II, conhecidos na literatura como *outliers* aditivos e inovadores, denotados por *AO* e *IO*, respectivamente. *Outliers* aditivos são observações discrepantes que afetam um único momento no tempo, por exemplo, um erro de digitação. Por outro lado, *outliers* inovadores afetam não somente uma observação, mas uma série de observações seguintes em um certo intervalo de tempo, por exemplo, uma enchente em um rio.

Na literatura, encontramos ainda questões referentes aos modelos geradores desses *outliers*, como o modelo de contaminação por mistura descritos por Denby e Martin (1979) e Bustos e Yohai (1986) e o modelo de contaminação paramétrico descrito por Hotta e Neves (1992) e Palma (1998).

3.1 Paramétrico

A seguir apresentamos os modelos paramétricos de contaminação para os processos ARFIMA (p, d, q) descritos por Hotta e Neves (1992) e Palma (1998).

Definição 3.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA (p, d, q) (ver Definição 2.3). O processo $\{Z_t\}_{t \in \mathbb{Z}}$ segue o *modelo paramétrico de contaminação* se é dado por

$$Z_t = X_t + \omega f(\mathcal{B})I_t(T), \quad (3.1)$$

onde $\omega \in \mathbb{R}$ representa a *magnitude* do *outlier*, $f(\mathcal{B})$ é uma *função de intervenção* e

$$I_t(T) = \begin{cases} 1, & \text{se } t = T, \\ 0, & \text{se } t \neq T. \end{cases} \quad (3.2)$$

é uma variável indicadora do tempo T em que ocorre o *outlier*.

Definição 3.2. Seja $f(\mathcal{B})$ a função de intervenção dada por $f(\mathcal{B}) = 1$, então podemos dizer que o processo $\{Z_t\}_{t \in \mathbb{Z}}$ está contaminado por um *outlier aditivo* (AO).

Definição 3.3. Seja $f(\mathcal{B})$ a função de intervenção dada por

$$f(\mathcal{B}) = \frac{\phi(\mathcal{B})}{\theta(\mathcal{B})}(1 - \mathcal{B}^d) \quad (3.3)$$

então podemos dizer que o processo $\{Z_t\}_{t \in \mathbb{Z}}$ está contaminado por um *outlier inovador* (IO).

Para o caso de um *outlier* aditivo, percebemos claramente pela equação (3.2) que apenas a T -ésima observação da série é afetada no processo $\{Z_t\}_{t \in \mathbb{Z}}$. Isto é principalmente caracterizado por erros grosseiros de digitação.

Por outro lado, no caso de um *outlier* inovador, percebemos que ocorrerá uma contaminação máxima na T -ésima observação, que transmite seu efeito para as observações seguintes através da função de intervenção representada pela equação (3.3). Esta perturbação é tipicamente observada em enchentes, greves, mudanças políticas ou econômicas, ondas de calor ou frio inesperadas.

O processo também pode seguir um modelo mais completo, onde existe a contaminação por m *outliers* de todos os tipos. A seguir, definimos este modelo como *modelo multi-paramétrico*.

Definição 3.4. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) (ver Definição 2.3). O processo $\{Z_t\}_{t \in \mathbb{Z}}$ segue o *modelo multi-paramétrico de contaminação* se é dado por

$$Z_t = X_t + \sum_{j=1}^m w_j f_j(\mathcal{B}) I_t(T_j), \quad (3.4)$$

onde w_j , $j = 1, \dots, m$, representa a *magnitude* do j -ésimo *outlier*, $I_t(T_j)$ é uma variável indicadora do j -ésimo momento em que ocorre a j -ésima perturbação dada pela equação (3.2), e $f_j(\mathcal{B})$ é a *função de intervenção* associada ao j -ésimo *outlier* e sua classificação como AO ou IO.

No modelo multi-paramétrico descrito na Definição 3.4 é fácil perceber que podemos ter somente *outliers* aditivos, somente *outliers* inovadores ou *outliers* aditivos e inovadores ao mesmo tempo.

3.2 Mistura

A seguir apresentamos os modelos de contaminação por mistura para os processos ARFIMA(p, d, q), descritos por Denby e Martin (1979) e Bustos e Yohai (1986).

A definição a seguir é um caso especial do modelo de contaminação descrito por Beran (1994) para *outliers* aditivos.

Definição 3.5. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação

$$Z_t := X_t + V_t, \quad (3.5)$$

onde $\{X_t\}_{t \in \mathbb{Z}}$ é um processo ARFIMA(p, d, q), com processo de inovação $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ (ver Definição 2.3). O processo $\{V_t\}_{t \in \mathbb{Z}}$ é constituído de variáveis aleatórias independentes e identicamente distribuídas com distribuição dada por

$$H_V = (1 - c)\delta_0 + cG, \quad (3.6)$$

onde $0 \leq c \leq 1$, δ_0 uma distribuição degenerada na origem, isto é, a esperança e a variância de uma variável aleatória com esta distribuição são ambas nulas e G uma distribuição arbitrária. Além disso, os processos $\{V_t\}_{t \in \mathbb{Z}}$ e $\{X_t\}_{t \in \mathbb{Z}}$ são independentes. Então, $\{Z_t\}_{t \in \mathbb{Z}}$ é dito ser um *processo com contaminação por mistura por outliers aditivos*.

A partir da equação (3.6) temos que:

- Se $c = 0$ não temos contaminação;
- Com probabilidade $1 - c$ o processo $\{Z_t\}_{t \in \mathbb{Z}}$ é igual ao processo $\{X_t\}_{t \in \mathbb{Z}}$;
- Com probabilidade c o processo $\{Z_t\}_{t \in \mathbb{Z}}$ é igual ao processo $\{X_t\}_{t \in \mathbb{Z}}$ adicionado de um erro $\{V_t\}_{t \in \mathbb{Z}}$. Quando este erro é adicionado, temos o comportamento de um *outlier*.

Quando a distribuição de probabilidade de G segue uma distribuição $N(0, \tau_V^2)$ temos que

- $\mathbb{E}(V_t) = 0$;
- $Var(V_t) = c\tau_V^2$;

As respectivas provas são encontradas em Bisognin (2007).

Definição 3.6. Seja $f_X(\cdot)$ a função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$ (ver equação 2.7) e $f_V(\omega) = \frac{c\tau_V^2}{2\pi}$ a função densidade espectral do processo $\{V_t\}_{t \in \mathbb{Z}}$. Então

$$f_Z(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \left[2 \operatorname{sen}\left(\frac{\omega}{2}\right) \right]^{-2d} + \frac{c\tau_V^2}{2\pi} \quad (3.7)$$

é a função densidade espectral do processo $\{Z_t\}_{t \in \mathbb{Z}}$

Definição 3.7. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) (ver Definição 2.3) que satisfaz a seguinte expressão

$$\phi(\mathcal{B})\nabla^d Z_t = \theta(\mathcal{B})\xi_t, \quad (3.8)$$

onde $\{\xi_t\}_{t \in \mathbb{Z}}$ é um processo que tem uma distribuição F_ξ com cauda pesada aproximadamente normal, isto é, uma distribuição normal contaminada, dada por

$$F_\xi = (1 - c)F_\varepsilon + cG, \quad (3.9)$$

onde $0 \leq c \leq 1$ é o parâmetro de mistura, F_ε uma distribuição $\mathcal{N}(0, \sigma_\varepsilon^2)$ e G uma distribuição arbitrária com uma dispersão superior a σ_ε^2 . Então, $\{Z_t\}_{t \in \mathbb{Z}}$ é dito ser um *processo* ARFIMA(p, d, q) *de acordo com o modelo IO*.

Quando $G = N(0, \tau_V^2)$, com $\tau_V^2 \geq \sigma_\varepsilon^2$ na equação (3.9), temos que

- $\mathbb{E}(\xi_t) = 0$;
- $\operatorname{Var}(\xi_t) = (1 - c)\sigma_\varepsilon^2 + c\tau_V^2$.

As respectivas provas são encontradas em Bisognin (2007).

Definição 3.8. Seja $f_X(\cdot)$ a função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$ (ver equação 2.7) e $f_\xi(\omega) = \frac{(1-c)\sigma_\varepsilon^2 + c\tau_V^2}{2\pi}$ a função densidade espectral do processo $\{\xi_t\}_{t \in \mathbb{Z}}$. Então

$$f_Z(\omega) = \left[\frac{(1 - c)\sigma_\varepsilon^2 + c\tau_V^2}{2\pi} \right] \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \left[2 \operatorname{sen}\left(\frac{\omega}{2}\right) \right]^{-2d} \quad (3.10)$$

é a função densidade espectral do processo $\{Z_t\}_{t \in \mathbb{Z}}$ com ruído $\{\xi_t\}_{t \in \mathbb{Z}}$.

Capítulo 4

Estimação e Identificação de *outliers*

Em estudos de séries temporais procuramos compreender e descrever o comportamento da série ao longo do tempo através de modelos. Em especial, neste trabalho estudamos os modelos com propriedade de longa dependência denominados ARFIMA (p, d, q) .

Na prática, raramente sabemos a priori a ordem dos polinômios $\phi(\cdot)$ e $\theta(\cdot)$, muito menos os valores de seus respectivos parâmetros, incluindo o parâmetro ou grau de diferenciação d (ver Definição 2.3). Para obtermos resultados a partir do modelo, é necessário estimar corretamente seus parâmetros.

Na literatura de séries temporais com propriedade de longa dependência existem diversos métodos para a estimação dos parâmetros. Em geral, os estimadores podem ser de três tipos, semi-paramétricos, paramétricos e não-paramétricos. Neste trabalho utilizaremos apenas o estimador paramétrico, que tem como característica estimar todos os parâmetros do modelo conjuntamente. O estimador *FT* utiliza o método da função de máxima verossimilhança aproximada descrita por Fox e Taqqu (1986) e Whittle (1951).

Quando existem *outliers* na série temporal, já sabemos que a estimativa dos parâmetros é afetada negativamente (veja por exemplo, Martin e Yohai 1986 e Peña 1990). Na literatura, encontramos diversos métodos e procedimentos iterativos para detecção de *outliers* principalmente para os modelos ARMA descritos inicialmente por Chang e Tiao (1983) e Tsay (1986) sendo estendido a partir do teste da razão de verossimilhança proposto por Fox (1972) para modelos AR puros. Em seguida Chang et al. (1988) estenderam este método para os modelos ARIMA. Como Tsay (1986) mencionou, existem diversas técnicas disponíveis para detectar apenas um *outlier*, mas existem complicações quando estendemos para diversas perturbações. Em geral, os métodos propostos exigem o conhecimento prévio das posições dos *outliers*, mas sabemos que na prática isso é muito difícil de acontecer. Enfrentando este problema, Choy (2001) propõe um procedimento para modelos AR(p)

e ARMA (p, q) denominado SODA, onde o *outlier* é detectado, estimado e eliminado em ciclos, até que não seja encontrado mais nenhuma perturbação. Para modelos de longa dependência sazonais, Bisognin (2007) estende o teste da razão de verossimilhança proposto por Fox (1972) para identificar e detectar *outliers* quando conhecemos suas posições na série.

Neste trabalho estendemos o método proposto por Choy (2001) para os modelos ARFIMA (p, d, q) e o comparamos com o método proposto por Bisognin (2007) para modelos SARFIMA $(p, d, q) \times (P, D, Q)_s$ quando conhecemos as posições dos *outliers*.

4.1 Estimação

A seguir apresentamos o método de estimação dos parâmetros por máxima verossimilhança aproximada para os casos onde não temos contaminação, onde a série é contaminada por *outliers* aditivos e onde a série é contaminada por *outliers* inovadores.

4.1.1 FT sem contaminação

Quando supomos que a série não está contaminada, o estimador de máxima verossimilhança aproximada descrito por Fox e Taquq (1986) e Whittle (1951) são utilizados para estimar todos os parâmetros do modelo conjuntamente. Fox e Taquq (1986) apresentam condições para que em uma sequência com dependência e estacionária, esse estimador seja consistente e tenha distribuição normal assintótica.

Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA (p, d, q) (ver Definição 2.3) com função densidade espectral $f_X(\cdot)$ dada pela equação (2.7) e $Q(\boldsymbol{\eta})$ a função dada por

$$Q(\boldsymbol{\eta}) = \int_{-\pi}^{\pi} \frac{I(w)}{f_X(w; \boldsymbol{\eta})} dw, \quad (4.1)$$

onde $f_X(\cdot; \boldsymbol{\eta})$ é a função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$ dada em (2.7) e $\boldsymbol{\eta}$ é o vetor de parâmetros que queremos estimar dado por

$$\boldsymbol{\eta} \equiv (\phi_1, \phi_2, \dots, \phi_p, d, \theta_1, \theta_2, \dots, \theta_q, \sigma_\varepsilon^2)$$

e $I(\cdot)$ é a função periodograma dada por

$$I(w) = \frac{1}{2\pi} \left[\hat{\gamma}_X(0) + 2 \sum_{k=1}^{n-1} \hat{\gamma}_X(k) \cos(wk) \right], \quad w \in [-\pi, \pi] \quad (4.2)$$

onde $\hat{\gamma}_X(\cdot)$ é a função de autocovariância amostral do processo $\{X_t\}_{t \in \mathbb{Z}}$.

Então, o estimador de máxima verossimilhança aproximado, denotado por FT , é o valor de $\boldsymbol{\eta}$ que minimiza a função $Q(\cdot)$. Para mais detalhes ver Whittle (1951).

O estimador de máxima verossimilhança aproximada é obtido computacionalmente na forma discreta, minimizando a função

$$\mathcal{L}_n(\boldsymbol{\eta}) = \frac{1}{2n} \sum_{j \in B} \left(\ln f_X(w_j, \boldsymbol{\eta}) + \frac{I(w_j)}{f_X(w_j, \boldsymbol{\eta})} \right), \quad (4.3)$$

que é uma aproximação da função $Q(\cdot)$ onde

$$w_j = \frac{2\pi j}{n}, j \in B = \left\{ 1, 2, \dots, \left\lfloor \frac{n-1}{2} \right\rfloor \right\} \quad (4.4)$$

são as frequências de Fourier.

4.1.2 FT contaminado por *outliers* aditivos

É fácil notar que quando consideramos o modelo de contaminação por mistura para *outliers* aditivos, utilizar o estimador FT sem contaminação não seria o ideal.

Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico contaminado por *outliers* aditivos seguindo o modelo de contaminação por mistura conforme Definição 3.5. O estimador FT contaminado por *outliers* aditivos denotado por $FTAO$ é dado pela equação (4.3) onde as frequências de Fourier são dadas pela equação (4.4), $f_Z(\cdot; \boldsymbol{\eta})$ é a função densidade espectral do processo $\{Z_t\}_{t \in \mathbb{Z}}$ dada em (3.7) e $\boldsymbol{\eta}$ é o vetor de parâmetros que queremos estimar dado por

$$\boldsymbol{\eta}_{AO} \equiv (\phi_1, \phi_2, \dots, \phi_p, d, \theta_1, \theta_2, \dots, \theta_q, \sigma_\varepsilon^2, c\tau_V^2)$$

e $I(\cdot)$ é a função periodograma dada pela equação (4.4).

4.1.3 FT contaminado por *outliers* inovadores

Assim como para o caso de *outliers* aditivos, é fácil notar que quando consideramos o modelo de contaminação por mistura para *outliers* inovadores, utilizar o estimador FT sem contaminação ou o $FTAO$ não seria o ideal.

Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico contaminado por *outliers* inovadores seguindo o modelo de contaminação por mistura conforme Definição 3.7. O estimador FT contaminado por *outliers* inovadores denotado por $FTIO$ é dado pela equação (4.3) onde as frequências de Fourier são dadas pela equação (4.4), $f_Z(\cdot; \boldsymbol{\eta})$ é a função densidade espectral do processo $\{Z_t\}_{t \in \mathbb{Z}}$ dada em (3.10) e $\boldsymbol{\eta}$ é o vetor de parâmetros que queremos estimar dado por

$$\boldsymbol{\eta}_{IO} \equiv (\phi_1, \phi_2, \dots, \phi_p, d, \theta_1, \theta_2, \dots, \theta_q, (1-c)\sigma_\varepsilon^2 + c\tau_V^2)$$

e $I(\cdot)$ é a função periodograma dada pela equação (4.4).

4.2 Identificação de *outliers*

Nesta seção apresentamos a metodologia proposta por Choy (2001) estendida para modelos ARFIMA(p, d, q) assim como a metodologia do teste da razão de verossimilhança descrito por Bisognin (2007). Ainda, apresentamos o método da função *tsoutlier* criado por Rob J. Hyndman disponível no pacote *forecast* do *software* R.

4.2.1 *Tsoutlier*

Para encontrar diversos *outliers* em uma série temporal, diversos autores sugerem algoritmos onde testamos se cada posição da série é realmente uma perturbação. Um exemplo disto é o próprio método proposto Choy (2001).

Para modelos de longa dependência, geralmente necessitamos de um tamanho amostral grande para uma boa estimativa dos parâmetros. Dado isso, é interessante utilizar um método mais simples para encontrar pelo menos as posições dos *outliers* verdadeiros e, então testar essas posições com um método mais robusto.

O procedimento *tsoutlier* implementado por Hyndman no pacote *forecast* do *software* R é apresentado a seguir com pequenos ajustes.

1. **Identificar sazonalidade**

O primeiro passo é verificar se existe sazonalidade na série temporal em estudo. Neste estudo trabalhamos com séries estacionárias com ausência de sazonalidade.

2. **Decompor a série em tendência e sazonalidade**

Se a série for sazonal, é realizada a estimação robusta dos componentes de tendência e sazonalidade por Loess (veja Cleveland et al, 1990) e então estes componentes são subtraídos da série original.

Se a série for não-sazonal, é realizada apenas a estimação robusta do componente de tendência por Loess (veja Cleveland et al, 1992) e então este componente é subtraído da série original.

Em ambos os casos, passamos a utilizar os resíduos da série resultante desta decomposição.

3. **Obter o primeiro e o terceiro quartil dos resíduos**

4. **Obter a amplitude interquartilica**

A amplitude interquartilica é obtida subtraindo o terceiro quartil do primeiro quartil.

5. Encontrar os limites para os *outliers*

A maneira de encontrar os limites para os resíduos *outliers* é a mesma utilizada para encontrar *outliers* no boxplot (veja Frigge, 1989). Se $Q1$ e $Q3$ são o primeiro quartil e o terceiro quartil, então as observações serão classificadas como *outliers* quando ultrapassarem os limites $Q1 - k(Q3 - Q1)$ e $Q3 + k(Q3 - Q1)$. Frigge (1989) faz uma comparação do impacto provocado por diferentes k na classificação de *outliers*. Neste trabalho utilizaremos $k = 1.5$.

6. Encontrar as posições das observações dos resíduos fora do limites

4.2.2 SODA

Spectrum-based outlier detection algorithm (SODA) proposto por Choy (2001) tem como objetivo estimar, detectar e remover *outliers* em modelos AR (p) ou ARMA (p, q). Neste trabalho estendemos este método para os modelos ARFIMA (p, d, q).

Definição 4.1. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA (p, d, q) (ver Definição 2.3) contaminado por *outliers* aditivos ou inovadores. Considere a estatística

$$D(I_Z, f_Z) = \int_{-\pi}^{\pi} \log f_Z(w) + \frac{I_Z(w)}{f_Z(w)} dw. \quad (4.5)$$

Então os estimadores de ω são dados por

$$\hat{\omega}(n, T) = \{\omega | \min\{D(I_{Z-\omega\delta}, f)\}\} \quad (4.6)$$

e para o caso *AO* e *IO*, respectivamente,

$$\hat{\omega}_{AO}(n, T) = \frac{\int_{-\pi}^{\pi} \sum_{t=1}^n Z_t e^{i\omega(t-T)} / f_Z(w) dw}{\int_{-\pi}^{\pi} 1 / f_Z(w) dw} \quad (4.7)$$

e

$$\hat{\omega}_{IO}(n, T) = \frac{\int_{-\pi}^{\pi} (\sum_{t=1}^n Z_t e^{i\omega(t-T)}) (\sum_{s=0}^{n-T} \psi_s e^{-i\omega s}) / f_Z(w) dw}{\int_{-\pi}^{\pi} |\sum_{t=0}^{n-T} \psi_t e^{i\omega s}|^2 / f_Z(w) dw}, \quad (4.8)$$

onde $f_Z(w)$ é a função densidade espectral do processo $\{Z_t\}_{t \in \mathbb{Z}}$ obtidos quando substituímos os parâmetros teóricos do modelo pelos seus respectivos valores estimados, e $\{\psi_k\}_{k \geq 0}$ são os coeficientes da representação média móvel infinita do processo $\{Z_t\}_{t \in \mathbb{Z}}$ contaminado por *outliers* aditivos ou inovadores.

O teorema a seguir pode ser encontrado em Choy (2001) para processos ARMA (p, q).

Teorema 4.1. *Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) (ver Definição 2.3) contaminado por outliers, então os estimadores baseados na aproximação de Whittle $\hat{\omega}_{AO}(n, T)$ e $\hat{\omega}_{IO}(n, T)$, definidos respectivamente pelas equações (4.7) e (4.8) tem distribuição:*

- $\hat{\omega}_{AO}(n, T) \rightarrow N(\omega, \sigma_{AO}^2)$, quando $n \rightarrow \infty$ e $1 \leq T \leq n$, onde

$$\sigma_{AO}^2 = \frac{(2\pi)^2}{\int_{-\pi}^{\pi} 1/f(w)dw};$$

- $\hat{\omega}_{IO}(n, T) \rightarrow N(\omega, \sigma_{IO}^2)$, quando $n \rightarrow \infty$ e $(n - T) \rightarrow \infty$, onde

$$\sigma_{IO}^2(m) = \frac{(2\pi)^2}{\int_{-\pi}^{\pi} |\sum_{t=0}^m \psi_t e^{iwt}|^2 / f(w)dw}.$$

Corolário 4.1. *Pelo Teorema 2.1 se $d < 0.5$ então o processo $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário e pode ser reescrito na forma de um modelo linear estacionário. Portanto pelo teorema 2.1 dado por Choy (2001), $\hat{\omega}_{AO}(n, T) \rightarrow N(\omega, \sigma_{AO}^2)$ quando $n \rightarrow \infty$, $1 \leq T \leq n$ e $\hat{\omega}_{IO}(n, T) \rightarrow N(\omega, \sigma_{IO}^2)$, quando $n \rightarrow \infty$ e $(n - T) \rightarrow \infty$.*

Assim, seguem as seguintes estatísticas:

$$\Gamma_{AO,T} = \frac{\hat{\omega}_{AO}(n, T)}{\sigma_{AO}} = \frac{\int_{-\pi}^{\pi} \sum_{t=1}^n Z_t e^{iw(t-T)} / f(w)dw}{2\pi \sqrt{\int_{-\pi}^{\pi} 1/f(w)dw}} \quad (4.9)$$

e

$$\Gamma_{IO,T} = \frac{\hat{\omega}_{IO}(n, T)}{\sigma_{IO}(n - T)} = \frac{\int_{-\pi}^{\pi} (\sum_{t=1}^n Z_t e^{iwt})(\sum_{s=0}^{n-T} \psi_s e^{-iw(s+T)}) / f(w)dw}{2\pi \sqrt{\int_{-\pi}^{\pi} |\sum_{t=0}^{n-T} \psi_t e^{iwt}|^2 / f(w)dw}} \quad (4.10)$$

Como ambas $\Gamma_{AO,T}$ e $\Gamma_{IO,T}$ tem distribuição assintótica $N(0, 1)$, podemos considerar essas estatísticas como critério para detecção de *outliers*. Para isso, é necessário estabelecer um ponto crítico C . Conforme sugerido por Chang et al. (1988) podemos considerar $C = 4$ para baixa sensibilidade, $C = 3.5$ para média sensibilidade e $C = 3$ para alta sensibilidade em modelos ARMA (p, q). Por simulação iremos encontrar um valor C para o caso dos processos ARFIMA (p, d, q).

Todo o procedimento teórico apresentado anteriormente é com base na existência de apenas um *outlier*, mas na prática também não sabemos quantos vamos encontrar. Para isso, o procedimento SODA otimizado pelo método *tsoutliers* visa detectar todos os possíveis *outliers* através de ciclos.

Etapa 1: Executar o método *tsoutlier* descrito na seção 4.2.1 para obter a posição das possíveis perturbações;

Etapa 2: Modelar a série como se não existissem *outliers* e estimar os seus respectivos parâmetros pelo método de Whittle;

Etapa 3: Calcular os valores de $\Gamma_{AO,T}$ e $\Gamma_{IO,T}$, onde T são as posições encontradas na Etapa 1, usando o modelo estimado na Etapa 2.

Se $\Gamma_T = |\Gamma_{AO,T}| > |\Gamma_{IO,T}| > C$, onde C é o ponto crítico definido anteriormente, então consideramos que existe um *outlier* aditivo na posição T .

Se $\Gamma_T = |\Gamma_{IO,T}| > |\Gamma_{AO,T}| > C$, onde C é o ponto crítico definido anteriormente, então consideramos que existe um *outlier* inovador na posição T_i .

Etapa 4: Eliminar os efeitos $\hat{w}_{AO}(n, T)$ ou $\hat{w}_{IO}(n, T)$ de acordo com a Etapa 3 e reestimar os parâmetros do modelo.

4.2.3 Teste da razão de verossimilhança

Inicialmente proposto por Fox (1972) para modelos AR puros, o teste da razão de verossimilhança tem como objetivo detectar e identificar *outliers* quando conhecemos suas posições na série temporal. A estimação da magnitude de perturbação w é feita através de regressão. Bisognin (2007) estende este método para processos SARFIMA $(p, d, q) \times (P, D, Q)_s$. O processo ARFIMA(p, d, q) é um caso particular do processo SARFIMA $(p, d, q) \times (P, D, Q)_s$, quando consideramos que não existe sazonalidade na série.

Consideramos a existência de um único *outlier* no tempo T e supomos que os parâmetros do processo $\{X_t\}_{t \in \mathbb{Z}}$ contaminado são conhecidos, assim como sua variância σ_ε^2 .

Para maiores detalher e demonstrações, ver Bisognin (2007).

1) *Outlier* Aditivo

Para estimar a magnitude de um *outlier* aditivo e testar a sua presença vamos considerar os resíduos do modelo ajustado.

Considere o modelo dado pela equação 3.4 quando temos apenas um *outlier* aditivo, ou seja, $m = 1$ e $f_j(\mathcal{B}) = 1$ Então podemos reescrevê-lo como

$$Z_t = \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} (1 - \mathcal{B})^{-d} \varepsilon_t + w I_t(T). \quad (4.11)$$

Assim, definimos

$$\mathcal{Y}_t = \varepsilon_t + w \pi(\mathcal{B}) I_t(T), \quad (4.12)$$

onde

$$\mathcal{Y}_t := \frac{\phi(\mathcal{B})}{\theta(\mathcal{B})}(1 - \mathcal{B})^d Z_t. \quad (4.13)$$

Assumindo que $1 \leq t \leq n$ a equação (4.12) pode ser escrita na forma de um sistema linear e como o processo $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ satisfaz as hipóteses (1) $\mathbb{E}(\varepsilon_j) = 0$, (2) $Var(\varepsilon_j) = \sigma_\varepsilon^2$ e (3) $Cov(\varepsilon_t, \varepsilon_k) = 0$, o estimador para o *outlier* aditivo através do método de mínimos quadrados é dado por

$$\hat{\omega}_{AO}(n, T) = \frac{\sum_{j=T}^n \pi_{j-T} \mathcal{Y}_j}{\sum_{j=T}^n \pi_{j-T}^2} = \rho_{AO}^2 \sum_{j=T}^n \pi_{j-T} \mathcal{Y}_j, \quad (4.14)$$

onde

$$\rho_{AO}^2 = \left(\sum_{j=T}^n \pi_{j-T}^2 \right)^{-1} \quad e \quad \pi_0 = 1.$$

Teorema 4.2. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) (ver definição 2.3) contaminado por *outlier* aditivo, então o estimador da perturbação $\hat{\omega}_{AO}(n, T)$ tem distribuição $N(\omega, \sigma_\varepsilon^2 \rho_{AO}^2)$.

A demonstração desse teorema pode ser encontrada em Bisognin (2007).

2) *Outlier* Inovador

Para estimar a magnitude de um *outlier* inovador e testar a sua presença vamos considerar os resíduos do modelo ajustado.

Considere o modelo dado pela equação (3.4) quando temos apenas um *outlier* inovador, ou seja, $m = 1$ e $f_j(\mathcal{B}) = \frac{\phi(\mathcal{B})}{\theta(\mathcal{B})}(1 - \mathcal{B}^d)$. Então podemos reescrevê-lo como

$$Z_t = \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})}(1 - \mathcal{B})^{-d}(\varepsilon_t + wI_t(T)). \quad (4.15)$$

Assim,

$$\mathcal{Y}_t = \varepsilon_t + wI_t(T), \quad (4.16)$$

onde

$$\mathcal{Y}_t := \frac{\phi(\mathcal{B})}{\theta(\mathcal{B})}(1 - \mathcal{B})^d Z_t. \quad (4.17)$$

Assumindo que $1 \leq t \leq n$ a equação (4.16) pode ser escrita na forma de um sistema linear e como o processo $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ satisfaz as hipóteses (1) $\mathbb{E}(\varepsilon_j) = 0$, (2) $Var(\varepsilon_j) = \sigma_\varepsilon^2$ e (3) $Cov(\varepsilon_t, \varepsilon_k) = 0$ o estimador para o *outlier* inovador através do método de mínimos quadrados é dado por

$$\hat{\omega}_{IO}(n, T) = \mathcal{Y}_T, \quad (4.18)$$

onde \mathcal{Y}_T é dado pela equação (4.16).

Teorema 4.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) (ver Definição 2.3) contaminado por *outlier* inovador, então o estimador da perturbação $\hat{\omega}_{IO}(n, T)$ tem distribuição $N(\omega, \sigma_\varepsilon^2)$.

A demonstração desse teorema pode ser encontrada em Bisognin (2007).

3) Teste de hipóteses

Para testar a presença de *outliers* pelo método de razão de verossimilhança consideramos conhecidos a posição do *outlier* utilizando o método *tsoutlier*, os parâmetros do modelo e sua variância σ_ε^2 . Então testar a significância do parâmetro ω é equivalente a testar

$$\begin{aligned} H_0 &: w = 0 \text{ nas equações (4.11) e (4.15),} \\ H_1^1 &: w \neq 0 \text{ na equação (4.11),} \\ H_1^2 &: w \neq 0 \text{ na equação (4.15).} \end{aligned}$$

Assim os testes da razão de verossimilhança para testar uma hipótese contra as outras são dados por

Teste 1: $H_0 : w = 0$ versus $H_1^1 : w \neq 0$, com estatística do teste para *AO* dada por

$$\lambda_{AO,T} = \frac{\hat{w}_{AO}}{\sigma_\varepsilon \rho_{AO}}. \quad (4.19)$$

Teste 2: $H_0 : w = 0$ versus $H_1^2 : w \neq 0$, com estatística do teste para *IO* dada por

$$\lambda_{IO,T} = \frac{\hat{w}_{IO}}{\sigma_\varepsilon}. \quad (4.20)$$

Sob hipótese nula H_0 as estatísticas $\lambda_{AO,T}$ e $\lambda_{IO,T}$, possuem distribuição $N(0, 1)$. Assim, podemos considerar $\lambda_{AO,T}$ e $\lambda_{IO,T}$ como critério de seleção de *outliers* quando a sua posição não é conhecida. Estabelecemos um ponto crítico C , da mesma forma que estabelecemos para o método *SODA* (ver Seção 4.2.2). Para este caso, também iremos encontrar o melhor C através de simulação.

Para cada possível *outlier* na posição T identificado pelo método *tsoutlier*, classificamos como *AO* se

$$|\lambda_{AO,T}| > |\lambda_{IO,T}| > C$$

e classificamos como *IO* se

$$|\lambda_{IO,T}| > |\lambda_{AO,T}| > C.$$

É importante observar que na prática os parâmetros do modelo também não são conhecidos, portanto precisamos estimá-los. Chang et al. (1988)

prova que as estatísticas são assintoticamente equivalentes às estatísticas de teste dada pelas equações (4.19) e (4.20) se usarmos o estimador de máxima verossimilhança.

Capítulo 5

Resultados por simulação

Neste capítulo apresentamos os resultados por simulação de Monte Carlo para o método *tsoutlier* definido na Seção 4.2.1, para os três métodos de estimação paramétrica, *FT*, *FTAO* e *FTIO* definidos nas Seções 4.1.1, 4.1.2 e 4.1.3 respectivamente, e para os métodos *SODA* e *Teste da razão de verossimilhança* definidos nas Seções 4.2.2 e 4.2.3 respectivamente.

As simulações realizadas neste trabalho foram realizadas através do software estatístico R.

Os *outliers* introduzidos nas séries foram

- **Somados** se o valor da série na posição T era positivo;
- **Subtraídos** se o valor da série na posição T era negativo.

Isto nos garante que os *outliers* estarão sempre destacados na série e nunca terão sua magnitude mascarada.

5.1 Resultados do método *tsoutlier*

Nesta seção apresentamos os resultados de identificação das posições dos *outliers* através do método *tsoutlier* apresentado na Seção 4.2.1 para processos ARFIMA (p, d, q) contaminados por *AO* e *IO* através do método de contaminação paramétrica (ver Seção 3.1).

As simulações são baseadas em 1000 replicações, com séries contaminadas pelo modelo de contaminação paramétrico com *outliers* do tipo *AO* e *IO* (ver Seção 3.1) nas posições pré-definidas $\{50, 100, 150, 200\}$. O objetivo é verificar que o método *tsoutliers* é eficaz na detecção das posições onde introduzimos perturbações nas séries.

São apresentados resultados com $d \in \{0.1, 0.3, 0.45\}$, $p \in \{0, 1\}$ com $\phi_1 = -0.6$, $q \in \{0, 1\}$ com $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$ para os casos *AO* e *IO*.

Através das Tabelas 5.1, 5.2 e 5.3 observamos que quando $p = 0$ e $q = 0$ temos uma detecção perfeita quando $d = 0.1$ e $d = 0.3$ tanto para os casos

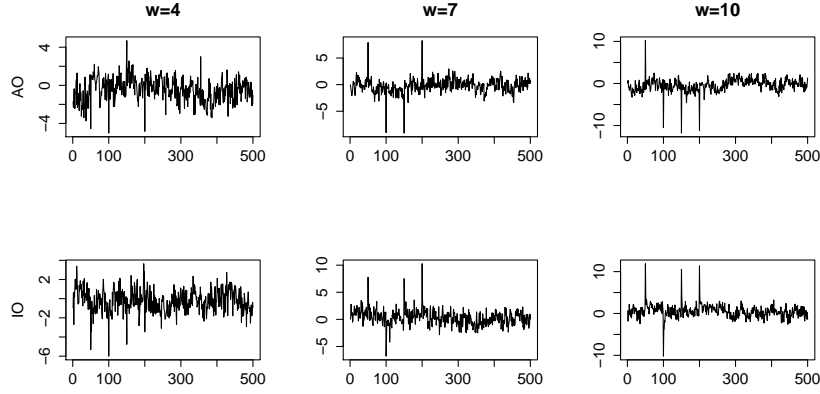


Figura 5.1: Processos ARFIMA (p, d, q) contaminados pelo método paramétrico nas posições $(50, 100, 150, 200)$, onde $p = 0$, $q = 0$, $d = 0.3$, $n = 500$ e $w \in \{4, 7, 10\}$.

AO quanto para os casos *IO* com magnitudes $w \in \{4, 7, 10\}$. Por outro lado, com $d = 0.45$ percebemos uma redução de eficiência quando $w = 4$ pesando mais para o caso *IO*.

Analisando as Tabelas 5.4, 5.5 e 5.6 observamos que quando $p = 1$ e $q = 0$ temos uma detecção perfeita quando $d = 0.1$ para os casos *AO*. Quando $d = 0.3$ temos uma melhoria na detecção de magnitude $w = 4$ para *IO* em relação a quando $d = 0.1$. Novamente, quando $w = 4$ notamos uma redução de eficiência para $w = 4$, mas uma detecção perfeita quando $w \in \{7, 10\}$.

Quando $q = 1$ e $q = 0$ observamos pelas Tabelas 5.7, 5.8 e 5.9 que se $d = 0.1$ temos uma boa detecção para $w \in \{4, 7, 10\}$ no caso *AO*, mas para o caso *IO* quando $w = 4$ percebemos uma queda na eficiência. Novamente, quando $d = 0.45$ notamos uma detecção baixa no caso *AO* quando $w \in \{4, 7\}$, mas quando $w = 10$ a detecção continua perfeita. No entanto, para o caso *IO*, vemos uma pequena redução de eficiência quando $w = 10$.

Em geral, notamos que o método *tsoutliers* funciona perfeitamente para detectar a posição dos *outliers* quando $w \in \{7, 10\}$. Para o caso de $w = 4$ é coerente a dificuldade de detecção, pois o efeito pequeno da perturbação pode ser mascarado pelo próprio comportamento da série (por exemplo, veja a Figura 5.1). Assim, consideramos o método válido para aperfeiçoar o desempenho dos métodos *SODA* e do *Teste da Razão de Verossimilhança*, já que sem ele seria necessário aplicar o método em todas as posições da série.

Tabela 5.1: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.1$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	1000	1000	1000	1000	1000	1000
75%	0	0	0	0	0	0
50%	0	0	0	0	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	1000	1000	1000	1000	1000	1000
75%	0	0	0	0	0	0
50%	0	0	0	0	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	1000	1000	1000	1000	1000	1000
75%	0	0	0	0	0	0
50%	0	0	0	0	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0

Tabela 5.2: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	995	1000	1000	985	1000	1000
75%	5	0	0	12	0	0
50%	0	0	0	3	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	998	1000	1000	988	1000	1000
75%	2	0	0	12	0	0
50%	0	0	0	0	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	1000	1000	1000	989	1000	1000
75%	0	0	0	11	0	0
50%	0	0	0	0	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0

Tabela 5.3: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.45$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	758	1000	1000	696	1000	1000
75%	187	0	0	239	0	0
50%	39	0	0	53	0	0
25%	15	0	0	12	0	0
0%	1	0	0	0	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	734	999	1000	673	1000	1000
75%	175	1	0	266	0	0
50%	74	0	0	54	0	0
25%	15	0	0	6	0	0
0%	2	0	0	1	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	653	1000	1000	580	1000	1000
75%	238	0	0	327	0	0
50%	92	0	0	74	0	0
25%	16	0	0	19	0	0
0%	1	0	0	0	0	0

Tabela 5.4: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.1$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	995	1000	1000	736	1000	1000
75%	5	0	0	227	0	0
50%	0	0	0	29	0	0
25%	0	0	0	7	0	0
0%	0	0	0	1	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	1000	1000	1000	840	1000	1000
75%	0	0	0	145	0	0
50%	0	0	0	14	0	0
25%	0	0	0	1	0	0
0%	0	0	0	0	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	1000	1000	1000	871	1000	1000
75%	0	0	0	118	0	0
50%	0	0	0	11	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0

Tabela 5.5: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.3$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	996	1000	1000	914	1000	1000
75%	4	0	0	85	0	0
50%	0	0	0	1	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	1000	1000	1000	947	1000	1000
75%	0	0	0	51	0	0
50%	0	0	0	2	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	1000	1000	1000	960	1000	1000
75%	0	0	0	39	0	0
50%	0	0	0	1	0	0
25%	0	0	0	0	0	0
0%	0	0	0	0	0	0

Tabela 5.6: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.45$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	891	1000	1000	919	1000	1000
75%	95	0	0	77	0	0
50%	13	0	0	4	0	0
25%	1	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	902	1000	1000	935	1000	1000
75%	78	0	0	57	0	0
50%	19	0	0	8	0	0
25%	1	0	0	0	0	0
0%	0	0	0	0	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	910	1000	1000	896	1000	1000
75%	67	0	0	100	0	0
50%	18	0	0	4	0	0
25%	5	0	0	0	0	0
0%	0	0	0	0	0	0

Tabela 5.7: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.1$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	978	1000	1000	702	1000	1000
75%	18	0	0	239	0	0
50%	4	0	0	57	0	0
25%	0	0	0	2	0	0
0%	0	0	0	0	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	997	1000	1000	769	1000	1000
75%	3	0	0	209	0	0
50%	0	0	0	20	0	0
25%	0	0	0	2	0	0
0%	0	0	0	0	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	1000	1000	1000	779	1000	1000
75%	0	0	0	207	0	0
50%	0	0	0	13	0	0
25%	0	0	0	1	0	0
0%	0	0	0	0	0	0

Tabela 5.8: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.3$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	696	1000	1000	259	986	1000
75%	219	0	0	359	13	0
50%	72	0	0	267	1	0
25%	11	0	0	97	0	0
0%	2	0	0	18	0	0
$n = 500$						
w	4	7	10	4	7	10
100%	660	1000	1000	281	996	1000
75%	259	0	0	375	4	0
50%	66	0	0	250	0	0
25%	15	0	0	84	0	0
0%	0	0	0	10	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	651	1000	1000	258	989	1000
75%	260	0	0	408	11	0
50%	72	0	0	255	0	0
25%	16	0	0	69	0	0
0%	1	0	0	10	0	0

Tabela 5.9: Número de séries que o método *tsoutliers* classifica corretamente (em %) para processos ARFIMA(p, d, q) contaminados por *AO* e *IO*, quando $d = 0.45$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

AO				IO		
$n = 250$						
w	4	7	10	4	7	10
100%	187	953	1000	39	617	966
75%	314	42	0	168	282	32
50%	306	5	0	309	85	1
25%	147	0	0	328	14	1
0%	46	0	0	156	2	0
$n = 500$						
w	4	7	10	4	7	10
100%	124	929	1000	31	683	988
75%	282	60	0	171	251	12
50%	332	11	0	320	60	0
25%	188	0	0	324	6	0
0%	74	0	0	154	0	0
$n = 1000$						
w	4	7	10	4	7	10
100%	84	887	1000	22	631	984
75%	247	91	0	124	317	15
50%	335	22	0	336	47	1
25%	251	0	0	369	5	0
0%	83	0	0	149	0	0

5.2 Resultados de Estimação pelos métodos FT , $FTAO$ e $FTIO$

Nesta seção apresentamos os resultados de estimação dos processos AR-FIMA (p, d, q) (ver Definição 2.3) contaminados por *outliers* aditivos e inovadores através do modelo de contaminação por mistura (ver seção 3.2).

Comparamos o desempenho dos estimadores paramétricos $FTAO$ e $FTIO$ com o estimador FT na presença de *outliers*.

As tabelas apresentadas nesta seção são referentes a $d = 0.3$ $p \in \{0, 1\}$, $q \in \{0, 1\}$, $\phi_1 = -0.6$, $\theta_1 = -0.6$, $w \in \{4, 7, 10\}$ e $n \in \{250, 500, 1000\}$. Os casos onde $d \in \{0.1, 0.45\}$ são apresentados no Apêndice A.

Os resultados são provenientes de 1000 replicações com séries contaminadas pelo modelo de contaminação por mistura para *outliers* dos tipos AO e IO com probabilidade de contaminação $c = 0.05$. Isto quer dizer que esperamos cerca de 50 *outliers* por série gerada.

Para o caso de séries contaminadas por *outliers* do tipo AO , as estimativas de d quando $p = 0 = q$ são sempre mais precisas para o método $FTAO$ mesmo com magnitude $w = 4$. Notamos também que o vício do método $FTAO$ diminui drasticamente conforme n aumenta, enquanto que para o método FT notamos uma leve diminuição. Quando $w = 10$ percebemos a ineficiência do método FT , enquanto que o método $FTAO$ ainda apresenta um vício muito pequeno, principalmente para $n = 1000$.

Quando $p = 1$ ou $q = 1$, as estimativas de d e ϕ_1 ou θ_1 seguem a mesma lógica de quando $p = 0 = q$, o método $FTAO$ possui menor vício para todas as magnitudes $w \in \{4, 7, 10\}$, assim como percebemos que esse vício diminui conforme n aumenta, e diminui muito pouco no método FT .

As estimativas de d , ϕ_1 ou θ_1 quando a série gerada é contaminada por *outliers* do tipo IO seguem a mesma lógica das contaminadas por AO , mas agora percebemos que ambos os métodos possuem um vício muito pequeno, independente de n e w , ainda que o método $FTIO$ seja minimamente mais preciso.

Tabela 5.10: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO				FT		
$n = 250$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.3354	0.3528	0.36	0.1976	0.1429	0.1068
vício	0.0354	0.0528	0.06	-0.1024	-0.1571	-0.1932
eqm	0.007	0.0106	0.0134	0.0136	0.0276	0.0396
variância	0.0058	0.0079	0.0099	0.0031	0.0029	0.0023
$n = 500$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.3455	0.3631	0.3747	0.2125	0.1504	0.1139
vício	0.0455	0.0631	0.0747	-0.0875	-0.1496	-0.1861
eqm	0.0045	0.008	0.0112	0.009	0.0238	0.0361
variância	0.0024	0.004	0.0056	0.0014	0.0014	0.0014
$n = 1000$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.343	0.367	0.3883	0.2152	0.1579	0.1226
vício	0.043	0.067	0.0883	-0.0848	-0.1421	-0.1774
eqm	0.0031	0.0061	0.0101	0.0079	0.0209	0.0321
variância	0.0012	0.0016	0.0023	0.0007	0.0007	0.0006

Tabela 5.11: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3, p = 1, \phi_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.3087	-0.6467	0.3081	-0.6683	0.3151	-0.6841	0.1996	-0.4198	0.1459	-0.3005	0.1134	-0.2233	
vício	0.0087	-0.0467	0.0081	-0.0683	0.0151	-0.0841	-0.1004	0.1802	-0.1541	0.2995	-0.1866	0.3767	
eqm	0.0073	0.0023	0.0124	0.0048	0.0154	0.0074	0.0154	0.0399	0.0284	0.0970	0.0391	0.1492	
variância	0.0072	0.0001	0.0124	0.0002	0.0152	0.0004	0.0053	0.0075	0.0046	0.0073	0.0043	0.0073	
$n = 500$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.3221	-0.6522	0.3210	-0.6796	0.3371	-0.6988	0.2182	-0.4356	0.1531	-0.3031	0.1136	-0.2243	
vício	0.0221	-0.0522	0.0210	-0.0796	0.0371	-0.0988	-0.0818	0.1644	-0.1469	0.2969	-0.1864	0.3757	
eqm	0.0039	0.0028	0.0083	0.0065	0.0125	0.0099	0.0091	0.0302	0.0242	0.0922	0.0371	0.1451	
variância	0.0034	0.0001	0.0079	0.0001	0.0112	0.0002	0.0024	0.0032	0.0026	0.0040	0.0024	0.0039	
$n = 1000$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.3257	-0.6560	0.3351	-0.6863	0.3503	-0.7077	0.2252	-0.4423	0.1611	-0.3105	0.1223	-0.2310	
vício	0.0257	-0.0560	0.0351	-0.0863	0.0503	-0.1077	-0.0748	0.1577	-0.1389	0.2895	-0.1777	0.3690	
eqm	0.0023	0.0032	0.0048	0.0075	0.0083	0.0117	0.0067	0.0264	0.0207	0.0859	0.0330	0.1383	
variância	0.0016	0.0000	0.0036	0.0001	0.0058	0.0001	0.0011	0.0016	0.0014	0.0021	0.0014	0.0022	

Tabela 5.12: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.2991	-0.7257	0.3223	-0.7553	0.3423	-0.7501	0.2523	-0.2745	0.2072	-0.1500	0.1684	-0.0862	
vício	-0.0009	-0.1257	0.0223	-0.1553	0.0423	-0.1501	-0.0477	0.3255	-0.0928	0.4500	-0.1316	0.5138	
eqm	0.0046	0.0163	0.0057	0.0257	0.0082	0.0237	0.0059	0.1061	0.0118	0.2027	0.0201	0.2642	
variância	0.0046	0.0005	0.0052	0.0016	0.0064	0.0012	0.0036	0.0002	0.0031	0.0002	0.0028	0.0002	
$n = 500$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.3027	-0.7403	0.3188	-0.7883	0.3384	-0.7880	0.2563	-0.2726	0.2220	-0.1336	0.1886	-0.0699	
vício	0.0027	-0.1403	0.0188	-0.1883	0.0384	-0.1880	-0.0437	0.3274	-0.0780	0.4664	-0.1114	0.5301	
eqm	0.0019	0.0200	0.0026	0.0364	0.0046	0.0380	0.0033	0.1073	0.0077	0.2176	0.0137	0.2812	
variância	0.0019	0.0003	0.0023	0.0009	0.0031	0.0026	0.0014	0.0001	0.0016	0.0001	0.0013	0.0001	
$n = 1000$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.3036	-0.7418	0.3125	-0.8241	0.3320	-0.8424	0.2666	-0.2631	0.2322	-0.1210	0.2012	-0.0571	
vício	0.0036	-0.1418	0.0125	-0.2241	0.0320	-0.2424	-0.0334	0.3369	-0.0678	0.4790	-0.0988	0.5429	
eqm	0.0009	0.0204	0.0013	0.0510	0.0023	0.0612	0.0018	0.1135	0.0053	0.2295	0.0104	0.2949	
variância	0,0009	0,0003	0.0011	0.0008	0.0013	0.0025	0,0007	0,0001	0,0007	0,0001	0,0006	0,0001	

Tabela 5.13: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO				FT		
$n = 250$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.3016	0.3009	0.2999	0.2878	0.2871	0.2861
vício	0.0016	0,0009	-1E-04	-0.0122	-0.0129	-0.0139
eqm	0,0001	0,0001	0,0001	0,0003	0,0003	0,0003
variância	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
$n = 500$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.2997	0.3013	0.3008	0.2923	0.2939	0.2935
vício	-3E-04	0.0013	0,0008	-0.0077	-0.0061	-0.0065
eqm	0,0000	0,0001	0,0000	0,0001	0,0001	0,0001
variância	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
$n = 1000$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.3011	0.3002	0.3007	0.2971	0.2962	0.2967
vício	0.0011	0,0002	0,0007	-0.0029	-0.0038	-0.0033
eqm	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
variância	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela 5.14: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3, p = 1, \phi_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.2956	-0.5933	0.2979	-0.5934	0.2975	-0.5939	0.2957	-0.5928	0.2976	-0.5930	0.2966	-0.5934	
vício	-0.0044	0.0067	-0.0021	0.0066	-0.0025	0.0061	-0.0043	0.0072	-0.0024	0.0070	-0.0034	0.0066	
eqm	0.0027	0.0002	0.0029	0.0002	0.0031	0.0002	0.0027	0.0002	0,003	0,0002	0.0031	0.0002	
variância	0.0027	0.0001	0.0029	0.0001	0.0031	0.0001	0.0027	0.0001	0,003	0,0001	0.0031	0.0001	
$n = 500$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.2962	-0.5961	0.2979	-0.5968	0.2962	-0.5971	0.2977	-0.5958	0.2970	-0.5967	0.2958	-0.5970	
vício	-0.0038	0.0039	-0.0021	0.0032	-0.0038	0.0029	-0.0023	0.0042	-0.0030	0.0033	-0.0042	0.0030	
eqm	0.0014	0.0001	0.0014	0.0001	0.0014	0.0001	0.0013	0.0001	0.0014	0.0001	0.0014	0.0001	
variância	0.0014	0.0001	0.0014	0.0001	0.0014	0.0001	0.0013	0.0001	0.0014	0.0001	0.0014	0.0001	
$n = 1000$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.2998	-0.5987	0.2998	-0.5986	0.3001	-0.5986	0.2996	-0.5985	0.2996	-0.5985	0.2999	-0.5984	
vício	-0.0002	0.0013	-0.0002	0.0014	0.0001	0.0014	-0.0004	0.0015	-0.0004	0.0015	-0.0001	0.0016	
eqm	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	
variância	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000	0,0006	0,0000	

Tabela 5.15: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.3, p = 1, \theta_1 = -0.6, n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.2968	-0.5945	0.2951	-0.5956	0.2931	-0.5944	0.2957	-0.5954	0.2944	-0.5964	0.2917	-0.5953	
vício	-0.0032	0.0055	-0.0049	0.0044	-0.0069	0.0056	-0.0043	0.0046	-0.0056	0.0036	-0.0083	0.0047	
eqm	0.0028	0.0001	0.0034	0.0001	0.0032	0.0001	0.0028	0.0001	0.0033	0.0001	0.0032	0.0001	
variância	0.0028	0.0001	0.0034	0.0001	0.0031	0.0001	0.0028	0.0001	0.0033	0.0001	0.0032	0.0001	
$n = 500$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.3004	-0.5974	0.2986	-0.5959	0.2995	-0.5973	0.2997	-0.5978	0.2977	-0.5963	0.2985	-0.5977	
vício	0.0004	0.0026	-0.0014	0.0041	-0.0005	0.0027	-0.0003	0.0022	-0.0023	0.0037	-0.0015	0.0023	
eqm	0.0015	0.0000	0.0014	0.0001	0.0014	0.0000	0.0015	0.0000	0.0014	0.0000	0.0014	0.0000	
variância	0.0015	0.0000	0.0014	0.0000	0.0014	0.0000	0.0015	0.0000	0.0014	0.0000	0.0014	0.0000	
$n = 1000$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.2996	-0.5983	0.2986	-0.5988	0.2990	-0.5972	0.2991	-0.5985	0.2984	-0.5990	0.2987	-0.5974	
vício	-0.0004	0.0017	-0.0014	0.0012	-0.0010	0.0028	-0.0009	0.0015	-0.0016	0.0010	-0.0013	0.0026	
eqm	0,0007	0,0000	0,0006	0,0000	0,0007	0,0000	0,0007	0,0000	0,0006	0,0000	0,0007	0,0000	
variância	0,0007	0,0000	0,0006	0,0000	0,0007	0,0000	0,0007	0,0000	0,0006	0,0000	0,0007	0,0000	

5.3 Resultados para os métodos *SODA* e *Teste da Razão de Verossimilhança*

Nesta seção apresentamos os resultados de estimação da magnitude w de perturbação para os métodos *SODA* e *Teste da Razão de Verossimilhança*, assim como propriedades de suas estatísticas de detecção $\Gamma_{\{AO, IO\}, T}$ e $\lambda_{\{AO, IO\}, T}$ com intuito de propor o melhor ponto critico para detecção de *outliers*. Os resultados são provenientes de processos ARFIMA (p, d, q) (ver Definição 2.3) contaminados por *outliers* aditivos e inovadores através do modelo de contaminação paramétrico (ver Seção 3.1).

Comparamos o desempenho dos estimadores da magnitude w de perturbação para os métodos *SODA* e *Teste da Razão de Verossimilhança*.

As tabelas apresentadas nesta seção são referentes a $d = 0.3$ $p \in \{0, 1\}$, $q \in \{0, 1\}$, $\phi_1 = -0.6$, $\theta_1 = -0.6$, $w \in \{7, 10\}$ e $n \in \{250, 500, 1000\}$.

Os resultados são provenientes de 1000 replicações com séries contaminadas pelo modelo de contaminação paramétrica na posição 100 para *outliers* dos tipos *AO* e *IO*.

Para o caso de contaminação por *outliers* do tipo *AO* onde $p = 0 = q$, $d = 0.3$ e $n = 250$, notamos que as estimativas de ambos estimadores são bastante precisas. Quando a magnitude da perturbação é 7 o estimador com menor vício é o w_{Reg} . Por outro lado, quando a magnitude de perturbação é 10, o estimador com menor vício é o w_{SODA} . Quando o n aumenta, percebemos que para ambas perturbações o estimador w_{Reg} possui menor vício.

Quando a contaminação é por *outliers* do tipo *IO* onde $p = 0 = q$ e $n = 250$, vemos um menor vício para o estimador w_{SODA} para a perturbação de magnitude 7 e o estimador w_{Reg} quando a perturbação é de magnitude 10. Quando aumentamos o n , percebemos que para ambas as magnitudes de perturbação o estimador com menor vício é o w_{SODA} .

Para os processos gerados com $p = 1$, $\phi_1 = -0.6$, $q = 0$, $d = 0.3$, $w \in \{7, 10\}$ e todos os tamanhos de amostra tanto para a contaminação por *AO* quanto para *IO*, o estimador de menor vício é o w_{Reg} . No entanto, percebemos que ambos os estimadores tem média muito próxima para o caso *AO*, mas para o caso *IO* temos uma notável piora do estimador w_{SODA} .

Para o caso onde $p = 0$, $q = 1$, $\theta_1 = -0.6$, $d = 0.3$ e $w = 7$ contaminado pelo tipo *AO*, notamos que para $n = 250$ e $n = 500$ o estimador com menor vício é o w_{SODA} . Para $n = 1000$ o estimador com menor vício é o w_{Reg} , apesar de ambas as estimativas serem muito boas. Para o caso onde $w = 10$, independente do tamanho amostral o estimador w_{SODA} apresenta o menor vício. Quando os processos são contaminado por *outliers* do tipo *IO*, percebemos uma drástica queda nas estimativas quando $w \in \{7, 10\}$ para todos tamanhos amostrais, ou seja, temos um vício muito grande neste caso. Por

outro lado o estimador w_{Reg} continua muito eficiente e com um vício pequeno para $w \in \{7, 10\}$ e todos tamanhos amostrais.

Em linhas gerais, o estimador w_{Reg} apresenta sinais de maior eficiência na estimação da magnitude de perturbação dos *outliers*. Notamos que quando o tamanho amostral aumenta, não temos uma melhor estimativa com ambos os estimadores. Outra característica interessante é que os dois métodos superestimam a magnitude w , o que corrobora com o fato de que quando esse *outlier* é podado a estimação dos parâmetros do modelo deixa de ser tão afetada.

Tabela 5.16: Estimação da magnitude da perturbação w , quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$.

AO					IO			
n=250								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	7.6846	7.6093	10.5922	10.5937	7.3872	7.7598	10.755	10.6299
vício	0.6846	0.6093	0.5922	0.5937	0.3872	0.7598	0.755	0.6299
eqm	0.9119	0.8648	0.9286	0.8766	1.0119	0.9653	0.98	0.8815
variância	0.4436	0.494	0.5784	0.5247	0.8655	0.3896	0.4104	0.4852
n=500								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	7.7198	7.5559	10.6578	10.5842	7.337	7.7982	10.1882	10.7283
vício	0.7198	0.5559	0.6578	0.5842	0.337	0.7982	0.1882	0.7283
eqm	0.9401	0.8322	0.8963	0.8714	0.7533	1.0279	0.7435	0.8809
variância	0.4223	0.5238	0.4641	0.5307	0.6423	0.3912	0.7109	0.3508
n=1000								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	7.8341	7.6432	10.7550	10.6299	7.3704	7.7712	10.2911	10.8583
vício	0.8341	0.6432	0.7550	0.6299	0.3704	0.7712	0.2911	0.8583
eqm	1.1137	0.9165	0.9800	0.8815	0.8427	0.9403	1.0217	1.1469
variância	0.4183	0.5033	0.4104	0.4852	0.7084	0.347	0.9404	0.4117

Tabela 5.17: Estimação da magnitude da perturbação w , quando $d = 0.3$, $p = 1$, $q = 0$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$.

AO					IO			
n=250								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	8.0921	7.9229	11.1261	10.9079	8.1159	7.6897	11.1816	10.7815
vício	1.0921	0.9229	1.1261	0.9079	1.1159	0.6897	1.1816	0.7815
eqm	2.5622	1.6219	2.7677	1.525	5.1712	1.3616	6.7906	1.4954
variância	1.3708	0.7710	1.5012	0.7015	3.9334	0.8875	5.4051	0.8864
n=500								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	8.1897	8.0461	11.1322	10.9683	8.1675	7.7407	11.1124	10.7219
vício	1.1897	1.0461	1.1322	0.9683	1.1675	0.7407	1.1124	0.7219
eqm	2.8159	2.0075	2.7723	1.8132	4.5462	1.4509	5.9289	1.3568
variância	1.4018	0.9141	1.492	0.8765	3.1893	0.9040	4.7007	0.8373
n=1000								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	8.1328	8.0189	11.1405	11.0047	8.0350	7.7010	11.2037	10.8041
vício	1.1328	1.0189	1.1405	1.0047	1.0350	0.7010	1.2037	0.8041
eqm	2.3627	1.8381	2.4519	1.7839	4.3343	1.4178	6.7069	1.6003
variância	1.0806	0.8009	1.1523	0.7752	3.2696	0.9282	5.2690	0.9558

Tabela 5.18: Estimação da magnitude da perturbação w , quando $d = 0.3$, $p = 0$, $q = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$.

AO					IO			
n=250								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	7.4846	7.6069	10.3872	10.6072	4.5188	7.7026	6.3825	10.6247
vício	0.4846	0.6069	0.3872	0.6072	-2.4812	0.7026	-3.6175	0.6247
eqm	0.6579	0.9923	0.537	0.9541	8.0832	1.1422	15.2846	1.2044
variância	0.4236	0.6246	0.3874	0.5859	1.9307	0.6498	2.2031	0.8158
n=500								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	7.5848	7.6314	10.4398	10.5828	4.6441	7.7889	6.401	10.6668
vício	0.5848	0.6314	0.4398	0.5828	-2.3559	0.7889	-3.599	0.6668
eqm	0.7549	1.0099	0.5896	0.9204	7.4307	1.3312	15.0625	1.1383
variância	0.4134	0.6118	0.3966	0.5813	1.8841	0.7103	2.1139	0.6951
n=1000								
	7		10		7		10	
	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}	\hat{w}_{SODA}	\hat{w}_{Reg}
média	7.6327	7.6099	10.5598	10.5986	4.6491	7.7348	6.4603	10.7798
vício	0.6327	0.6099	0.5598	0.5986	-2.3509	0.7348	-3.5397	0.7798
eqm	0.8009	0.9301	0.6955	0.9117	7.4589	1.2052	15.0568	1.3075
variância	0.4011	0.5587	0.3825	0.554	1.9359	0.6666	2.5326	0.7009

Para a detecção das posições dos possíveis *outliers* foi possível verificar que o método *tsoutliers* funciona adequadamente quando a série é estacionária. Quando d é muito próximo de 0.5 observamos que a série começa a apresentar sinais de não estacionariedade, o que acaba dificultando a detecção das possíveis posições.

Após detectar as possíveis posições, constatamos que ambos os estimadores de magnitude de perturbação w_{SODA} e w_{Reg} funcionam adequadamente. Como sabemos que o método *tsoutliers* identifica em geral sempre mais posições do que as verdadeiras, utilizamos as estatísticas de teste dos métodos *SODA* e *Teste da Razão de Verossimilhança* dadas por Γ e λ respectivamente, para identificarmos apenas os *outliers* verdadeiros.

Através das Tabelas 5.19, 5.20 e 5.21, percebemos claramente a necessidade de diferenciar o ponto crítico C de acordo com a magnitude w dos *outliers* e do método que está sendo utilizado.

Para o caso onde $d = 0.3$, $p = 0 = q$ e $w = 7$, independente do tamanho amostral e do tipo de *outlier*, um $C = 2.50$ já seria suficiente para detectarmos todos os *outliers* verdadeiros através do método *SODA*. Por outro lado, para o método do *Teste da Razão de Verossimilhança* um $C = 5.50$ seria suficiente. Se olharmos conjuntamente o mínimo e a média das estatísticas de detecção, percebemos que o ponto de corte para a estatística Γ poderia ser também $C = 3$ e para a estatística λ poderia ser $C = 6.5$ para que em média, todos os *outliers* verdadeiros sejam selecionados. No caso de $w = 10$, para o método *SODA* um $C = 3.3$ seria suficiente para detectar todas as perturbações, enquanto que para o método *Teste da Razão de Verossimilhança* o ideal seria $C = 7.6$.

Quando $d = 0.3$, $p = 1$, $q = 0$, $\phi_1 = -0.6$ e $w = 7$ independente do tamanho amostral e do tipo de *outlier*, um $C = 2$ já seria suficiente para detectarmos todos os *outliers* verdadeiros através do método *SODA*. Por outro lado, para o método do *Teste da Razão de Verossimilhança* um $C = 3.3$ seria suficiente. Se olharmos conjuntamente o mínimo e a média das estatísticas de detecção, percebemos que o ponto de corte para a estatística Γ poderia ser também $C = 3.5$ e para a estatística λ poderia ser $C = 5.5$ para que em média, todos os *outliers* verdadeiros sejam selecionados. No caso de $w = 10$, para o método *SODA* um $C = 3.4$ seria suficiente para detectar todas as perturbações, enquanto que para o método *Teste da Razão de Verossimilhança* o ideal seria $C = 5.2$.

Quando $d = 0.3$, $p = 0$, $q = 1$, $\theta_1 = -0.6$ e $w = 7$ independente do tamanho amostral e do tipo de *outlier*, para o método *Teste da Razão de Verossimilhança* um $C = 4.2$ já seria suficiente para selecionar todos os *outliers* verdadeiros. Para o método *SODA*, no caso de *outliers* do tipo *AO* um $C = 2$ já seria suficiente. No caso de *outliers* do tipo *IO*, o método *SODA* não funcionou conforme o esperado nem na estimativa da magnitude de perturbação w , portanto o mínimo de detecção seria um C muito baixo.

Mas, se considerarmos a média, poderíamos sugerir um $C = 4$. No caso de $w = 10$, para o método *SODA*, considerando seu problema de estimação no caso *IO*, um $C = 3.7$ seria suficiente para detectar todas as perturbações, enquanto que para o método *Teste da Razão de Verossimilhança* o ideal seria $C = 7$.

Em linhas gerais, como ambas as estatísticas de detecção Γ e λ possuem distribuição assintótica $N(0,1)$, temos que ter cuidados apenas com valores de C abaixo de 3. Levando isso em conta percebemos que o método *SODA* pode classificar como *outliers* casos falsos.

Uma proposta para o método *SODA* seria utilizar $C = 2$ para *outliers* de magnitude média e $C = 3$ para magnitudes grandes. Para o método *Teste da Razão de Verossimilhança* sugerimos $C = 4$ para magnitudes médias e grandes, pois em média os valores da estatística λ são muito altos, sendo de fácil detecção.

Tabela 5.19: Estatísticas descritivas das estatísticas de detecção de *outliers* quando $d = 0.3$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$.

AO					IO			
n=250								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.44	5.47	3.71	7.80	2.13	5.82	3.35	7.79
média	3.19	7.18	4.71	9.13	3.85	6.93	5.43	8.87
n=500								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.45	5.57	3.49	7.63	2.20	6.16	3.33	8.56
média	3.12	7.52	4.4	9.99	3.99	7.38	5.49	9.69
n=1000								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.51	5.67	3.71	8.76	2.16	6.38	3.37	9.09
média	3.14	7.81	4.34	10.57	3.86	7.57	5.44	10.23

Tabela 5.20: Estatísticas descritivas das estatísticas de detecção de *outliers* quando $d = 0.3$, $p = 1$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$.

AO					IO			
n=250								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.05	4.29	3.58	5.23	2.37	3.69	3.44	5.43
média	3.72	6.16	5.60	7.97	6.76	5.55	9.13	7.12
n=500								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.57	3.36	3.43	5.95	2.57	3.36	3.43	5.95
média	6.44	5.89	8.75	7.75	6.44	5.89	8.75	7.75
n=1000								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.41	4.69	3.73	6.59	2.30	3.98	3.43	5.94
média	3.53	6.60	4.89	8.88	6.28	5.99	9.10	8.18

Tabela 5.21: Estatísticas descritivas das estatísticas de detecção de *outliers* quando $d = 0.3$, $p = 0$, $q = 1$, $n \in \{250, 500, 1000\}$ e $w \in \{7, 10\}$.

AO					IO			
n=250								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.05	4.22	3.87	6.73	0.05	4.30	0.06	5.85
média	3.39	6.31	5.19	8.24	5.69	5.95	4.76	7.55
n=500								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.48	4.76	3.75	7.1	0.54	4.61	0.78	6.64
média	3.29	6.61	4.71	8.84	4.07	6.34	8.50	8.27
n=1000								
	7		10		7		10	
	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$	$\hat{\Gamma}$	$\hat{\lambda}$
min	2.46	4.90	3.81	7.52	0.04	4.19	0.83	7.00
média	3.25	6.73	4.57	9.17	4.56	6.45	5.14	8.77

5.4 Exemplo de procedimento completo

Nesta seção, apresentamos resultados do ciclo completo dos métodos *SODA* e *Teste da Razão de Verossimilhança*, onde temos como objetivo estimar, identificar e corrigir os *outliers* com a finalidade de estimar corretamente os parâmetros do modelo.

Para comparar os resultados, utilizaremos a mesma série gerada com $d = 0.3$, $p = 0 = q$ e *outliers* do tipo *AO* contaminados nas posições (50, 100, 200, 500) com $w = 10$ em uma série de tamanho 1000.

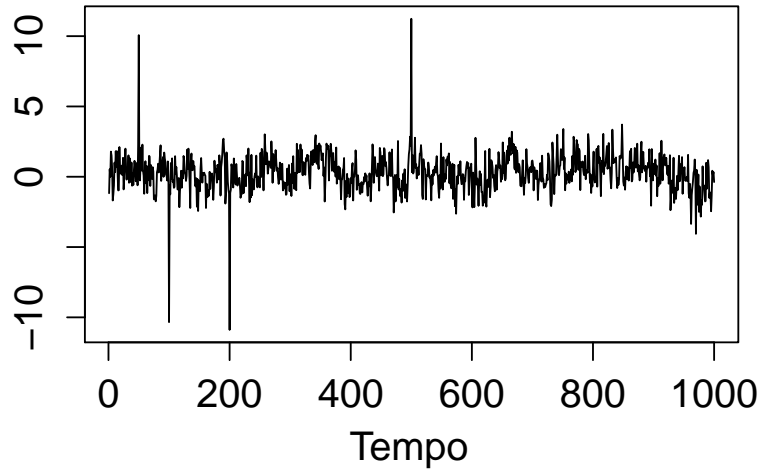


Figura 5.2: Processo ARFIMA (p, d, q) contaminados pelo método paramétrico nas posições (50, 100, 200, 500), onde $p = 0 = q$, $d = 0.3$, $n = 1000$ e $w = 10$.

Para a série apresentada na Figura 5.2 percebemos claramente os *outliers* que estão contaminando o processo original.

As estimativas de d para a série contaminada são apresentadas na Tabela 5.22. Em seguida apresentamos as posições resultantes do método *tsoutlier* e os resultados passo a passo dos métodos *SODA* e *Teste da Razão de Verossimilhança*.

Tabela 5.22: Estimativas do parâmetro d para o processo ARFIMA (p, d, q) contaminado por *AO* quando $p = 0 = q$ e $d = 0.3$.

	<i>FT</i>	<i>FTAO</i>
\hat{d}	0.2558	0.3287

Para ambos os métodos a seguir, a função *tsoutlier* propõe as mesmas posições para a série em estudo. São elas (50, 100, 200, 500, 832, 848, 962, 970).

Percebemos que o método *tsoutliers* foi capaz de selecionar todas as posições verdadeiras.

1) Método *SODA*

A partir do modelo ARFIMA (p, d, q) onde $p = 0 = q$, estimamos o parâmetro d e obtemos a estimativa $\hat{d} = 0.3287$ para o método *FTAO* apresentado na Tabela 5.22. Assim, utilizando as posições encontradas pelo método *tsoutlier* podemos executar o procedimento. Os resultados são apresentados na Tabela 5.23.

Tabela 5.23: Estimativas da magnitude de perturbação w e sua estatística de detecção Γ para o processo ARFIMA (p, d, q) contaminado por *AO* quando $p = 0 = q$ e $\hat{d} = 0.3287$.

Posição	\hat{w}	$\hat{\Gamma}$
50	9.3062	4.0951
100	-10.1249	4.4553
200	-10.1189	4.4527
500	9.9958	4.3985
832	2.4694	1.0866
848	2.4744	1.0888
962	-2.4370	1.0724
970	-4.0221	1.7699

Para as posições sugeridas, observamos que a estatística de detecção Γ seleciona perfeitamente as posições verdadeiras e estima as respectivas magnitudes de perturbação com bastante precisão. Aplicando o ponto de corte sugerido na Seção anterior $C = 3$ percebemos que apenas as posições verdadeiras (50, 100, 200, 500) são selecionadas como *outliers*.

Quando retiramos as magnitudes de perturbações estimadas \hat{w} para suas respectivas posições, re-estimamos os parâmetros do modelo. As novas estimativas são apresentadas na Tabela 5.24.

Tabela 5.24: Estimativas do parâmetro d para o processo ARFIMA (p, d, q) contaminado por *AO* quando $p = 0 = q$ e $d = 0.3$ após aplicar o método *SODA*.

	<i>FT</i>	<i>FTAO</i>
\hat{d}	0.2867	0.2980

É possível verificar que ambos os estimadores *FT* e *FTAO* apresentaram melhoras significativas na estimação do parâmetro d do modelo e, ainda é possível notar que o estimador *FTAO* possui melhor estimativa.

1) Método *Teste da Razão de Verossimilhança*

A partir do modelo ARFIMA (p, d, q) onde $p = 0 = q$, estimamos o parâmetro d e obtemos a estimativa $\hat{d} = 0.3287$ para o método *FTAO* apresentado na Tabela 5.22. Assim, utilizando as posições encontradas pelo método *tsoutlier* podemos executar o procedimento. Os resultados são apresentados na Tabela 5.25.

Tabela 5.25: Estimativas da magnitude de perturbação w e sua estatística de detecção λ para o processo ARFIMA (p, d, q) contaminado por *AO* quando $p = 0 = q$ e $\hat{d} = 0.3287$.

Posição	\hat{w}	$\hat{\lambda}$
50	9.3998	8.1695
100	-10.2206	8.8829
200	-10.2945	8.9471
500	10.1177	8.7935
832	2.5304	2.1992
848	2.6308	2.2865
962	-2.5552	2.2207
970	-3.9499	3.4327

Para as posições sugeridas, observamos que a estatística de detecção λ seleciona perfeitamente as posições verdadeiras e estima as respectivas magnitudes de perturbação com bastante precisão. Aplicando o ponto de corte sugerido na Seção anterior $C = 4$ percebemos que apenas as posições verdadeiras (50, 100, 200, 500) são selecionadas como *outliers*.

Quando retiramos as magnitudes de perturbações estimadas \hat{w} para suas respectivas posições, re-estimamos os parâmetros do modelo. As novas estimativas são apresentadas na Tabela 5.26.

Tabela 5.26: Estimativas do parâmetro d para o processo ARFIMA (p, d, q) contaminado por *AO* quando $p = 0 = q$ e $d = 0.3$ após aplicar o método *Teste da Razão de Verossimilhança*.

	<i>FT</i>	<i>FTAO</i>
\hat{d}	0.2869	0.2981

É possível verificar que ambos os estimadores *FT* e *FTAO* apresentaram melhoras significativas na estimação do parâmetro d do modelo e, ainda é possível notar que o estimador *FTAO* possui melhor estimativa.

Capítulo 6

Conclusão

Neste trabalho descrevemos importantes propriedades dos processos ARFIMA (p, d, q) , que tem como objetivo modelar séries com longa dependência, ou seja, quando séries são caracterizadas pela presença de correlação entre observações separadas por um longo período de tempo. Para este processo destacamos os Teoremas 2.1 e 2.2, onde apresentamos propriedades como estacionariedade, causalidade, invertibilidade, sua função de autocovariância, autocorrelação e função densidade espectral.

Descrevemos o modelo de contaminação paramétrico descrito por Hotta e Neves (1992) e Palma (1998) para os processos ARFIMA (p, d, q) , assim como o modelo de contaminação por mistura descrito por Denby e Martin (1979) e Bustos e Yohai (1986). Para o modelo de contaminação paramétrico destacamos a função de intervenção para que altera o efeito de *outliers* do tipo *AO* e *IO*, e para o modelo de contaminação por mistura destacamos a sua função densidade espectral para o caso de *outliers* do tipo *AO* e *IO*.

Descrito por Bisognin (2007), utilizamos os estimadores *FTAO* e *FTIO* que consideram o efeito dos *outliers* para realizar uma estimação mais precisa em relação ao estimador *FT*. Em todos os estimadores é utilizado como base o método de máxima verossimilhança aproximada descrita por Fox e Taqqu (1986) e Whittle (1951).

Estendemos o método *SODA* introduzido por Choy (2001) para identificação, estimação e correção de *outliers* de processos ARMA (p, q) para processos ARFIMA (p, d, q) . Também utilizamos o método do *Teste da Razão de Verossimilhança* descrito por Bisognin (2007) para processos SARFIMA $(p, d, q) \times (P, D, Q)_S$ onde a magnitude de perturbação é estimada através de regressão, e sugerimos um procedimento para que esse método seja comparado com método *SODA*. Para otimizar os algoritmos, utilizamos o método *tsoutlier* para detectar as possíveis posições dos *outliers* e mostramos através de simulações de Monte Carlo que este método é eficaz.

Obtemos diversos resultados para os métodos em estudo através de simulações de Monte Carlo. Verificamos a que o método de estimação *FTAO* e *FTIO* são mais eficientes na presença de *outliers* do tipo *AO* e *IO*, com

séries geradas variando $n \in \{250, 500, 1000\}$, $d \in \{0.1, 0.3, 0.45\}$, $p \in \{0, 1\}$, $q \in \{0, 1\}$, $\phi_1 = -0.6$, $\theta_1 = -0.6$ e perturbações de magnitude $w \in \{4, 7, 10\}$ com probabilidade de ocorrência $p = 0.05$.

Para testar o método *tsoutliers* utilizamos séries geradas variando $n \in \{250, 500, 1000\}$, $d \in \{0.1, 0.3, 0.45\}$, $p \in \{0, 1\}$, $q \in \{0, 1\}$, $\phi_1 = -0.6$, $\theta_1 = -0.6$ e perturbações de magnitude $w \in \{4, 7, 10\}$ nas posições $(50, 100, 150, 200)$. Como resultado temos que o método funciona perfeitamente para $w \in \{7, 10\}$. Quando $w = 4$ temos uma detecção menor das posições corretas, mas é coerente com o fato de que a magnitude pequena possa ser confundida com o efeito natural do processo, dificultando sua detecção.

Para testar os métodos de estimação da magnitude dos *outliers* através do *SODA* e do *Teste da Razão de Verossimilhança*, utilizamos para gerar as séries $n \in \{250, 500, 1000\}$, $d = 0.3$, $p \in \{0, 1\}$, $q \in \{0, 1\}$, $\phi_1 = -0.6$, $\theta_1 = -0.6$ e perturbações de magnitude $w \in \{7, 10\}$ na posição 100. Como resultado, o estimador w_{Reg} apresenta sinais de maior eficiência na estimação da magnitude de perturbação dos *outliers* tanto *AO* quanto *IO*. Notamos que quando o tamanho amostral aumenta, não temos uma melhor estimativa com ambos os estimadores. Outra característica interessante é que os dois métodos superestimam a magnitude w , o que corrobora com o fato de que quando esse *outlier* é podado a estimação dos parâmetros do modelo deixa de ser tão afetada.

Com intuito de sugerir pontos de corte adequados para as estatísticas de detecção dos métodos *SODA* e *Teste da Razão de Verossimilhança*, denotadas por Γ e λ , analisamos os valores de mínimo e média para as mesmas séries geradas para avaliar a estimação da magnitude dos *outliers*. Como resultado, sugerimos que para o método *SODA* utilizar $C = 2$ para *outliers* de magnitude média e $C = 3$ para magnitudes grandes. Para o método *Teste da Razão de Verossimilhança* sugerimos $C = 4$ para magnitudes médias e grandes, pois em média os valores da estatística λ são muito altos, sendo de fácil detecção.

Por fim, podemos observar através de uma única série que ambos os métodos *SODA* e *Teste da Razão de Verossimilhança* funcionam para identificar, estimar e corrigir *outliers* em processos ARFIMA (p, d, q) , resultando em uma estimação mais precisa para os parâmetros do modelo.

É importante salientar que não foram consideradas todas as possíveis variações nos parâmetros do modelo. Não apresentamos resultados para o processo ARFIMA (p, d, q) para o caso onde $p = 1$ e $q = 1$. Além disso, todos os resultados são baseados na suposição de estacionariedade do processo. Sugere-se para trabalhos futuros analisar os métodos descritos para diferentes valores de p , d e q . Também sugere-se estender os procedimentos deste trabalho para demais processos com longa dependência.

Referências

- [1] Baillie, R.T. (1996). “Long Memory Processes and Fractional Integration in Econometrics”. *Journal of Econometrics*, Vol. **73**, pp. 5-59.
- [2] Barnett, V. (1984). *Outliers in statistical data*. New York: John Wiley.
- [3] Beran, J. (1994). *Statistics for Long-Memory Processes*. New York: Chapman & Hall.
- [4] Bisognin, C. (2007). “Estimação e previsão em processos SARFIMA $(p,d,q) \times (P,D,Q)_s$ na presença de outliers”. 216 f. Tese (Doutorado em Matemática) – Instituto de Matemática. Universidade Federal do Rio Grande do Sul, Porto Alegre.
- [5] Bisognin, C. e S.R.C. Lopes (2005). “Processos Fracionários Generalizados”. *Anais da 11 Escola de Séries Temporais e Econometria*, Vol **1**, pp. 111-111.
- [6] Bisognin, C. e S.R.C. Lopes (2007). “Estimating and Forecasting the Long Memory Parameter in the Presence of Periodicity”. *Journal of Forecasting*, Vol. **26**(6), pp. 405-427.
- [7] Box, G.E.P., G.M. Jenkins e G.C. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. New Jersey: Prentice Hall.
- [8] Box, G.E.P. e G.C. Tiao (1975). “Intervention Analysis with Applications to Economic and Environmental Problems”. *Journal of the American Statistical Association*, Vol. **70**, pp. 70-79.
- [9] Brockwell, P.J. e R.A. Davis (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- [10] Bustos, O.H. e V.J. Yohai (1986). “Robust Estimates for ARMA Models”. *Journal of the American Statistical Association*, Vol. **81**, pp. 155-168.

- [11] Chang, I. e G.C. Tiao (1983). "Estimation of time series parameters in the presence of outliers". Technical Report 8, Statistics Research Center, University of Chicago.
- [12] Chang, I.H., G.C. Tiao e C. Cheng (1988). "Estimation of Time Series Parameters in the Presence of Outliers". *Technometrics*, Vol. **30**, pp. 193-204.
- [13] Chen, C., Liu, L. M. (1993). "Joint estimation of model parameters and outlier effects in time series". *Journal of the American Statistical Association*, 88(421), 284-297.
- [14] Choy, K. (2001). "Outlier detection for stationary time series". *Journal of Statistical Planning and Inference*, 99(2), 111-127
- [15] Chung, C-F. (1996). "Estimating a Generalized Long Memory Process". *Journal of Econometrics*, Vol. **73**, pp. 237-259.
- [16] Collet, J., D. Guégan e P. Valdes-Sosa (2003). "How Shall We Determine the Number and the Location of the Gegenbauer Frequencies? An Empirical Approach". Note de Recherche IDHE-MORA n° 2003-09.
- [17] Denby, L. e R.D. Martin (1979). "Robust Estimation of the First-Order Autoregressive Parameter". *Journal of the American Statistical Association*, Vol. **74**, pp. 140-146.
- [18] Doukhan, P., G. Oppenheim e M.S. Taqqu (2003). *Theory and Applications of Long-Range Dependence*. Boston: Birkhäuser.
- [19] Frigge, Michael; Hoaglin, David C.; Iglewicz, Boris (February 1989). "Some Implementations of the Boxplot". *The American Statistician* 43 (1): 50-54.
- [20] Fox, A.J. (1972). "Outliers in Time Series". *Journal of the Royal Statistical Society*, Vol. **B-43**, pp. 350-363.
- [21] Fox, R. e M.S. Taqqu (1986). "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series". *The Annals of Statistics*, Vol. **14**, pp. 517-532.
- [22] Geweke, J. e S. Porter-Hudak (1983). "The Estimation and Application of Long Memory Time Series Model". *Journal of Time Series Analysis*, Vol. **4**(4), pp. 221-238.
- [23] Granger, C.W.J. e R. Joyeux (1980). "An Introduction to Long Memory Time Series Models and Fractional Differencing". *Journal of Time Series Analysis*, Vol. **1**(1), pp. 15-29.

- [24] Gray, H. L., N-F. Zhang e W.A. Woodward (1989). "On Generalized Fractional Processes". *Journal of Time Series Analysis*, Vol. **10**(3), pp. 233-257.
- [25] Hosking, J.R.M. (1981). "Fractional Differencing". *Biometrika*, Vol. **68**(1), pp. 165-176.
- [26] Hosking, J.R.M. (1984). "Modelling Persistence in Hydrological Time Series Using Fractional Differencing". *Water Resources Research*, Vol. **20**(12), pp. 1898-1908.
- [27] Hurst, H.E. (1951). "Long-term Storage Capacity of Reservoirs". *Transactions of American Society in Civil Engineers*, Vol. **116**, pp. 165-167.
- [28] Lopes, S.R.C., B.P. Olbermann e V.A. Reisen (2004). "A Comparison of Estimation in Non-Stationary ARFIMA Processes". *Journal of Statistical Computation and Simulation*, Vol. **74**(5), pp. 339-347.
- [29] Reisen V.A., S.R.C. Lopes e M.R. Sena (2002). "Error and Model Misspecification in ARFIMA Processes". *Brazilian Review Of Econometrics*, Vol. **abf 21**(1), pp. 101-135.
- [30] R. B. Cleveland, W. S. Cleveland, J.E. McRae, and I. Terpenning (1990) "STL: A Seasonal-Trend Decomposition Procedure Based on Loess". *Journal of Official Statistics*, 6, 3-73.
- [31] Morettin P.A. e C.M.C. Toloi (2004). *Análise de Séries Temporais*. São Paulo: Edgard Blücher.
- [32] Palma, J.A. (1998). *Outliers em Séries Temporais - Uma Abordagem no Domínio dos Modelos ARMA*. Dissertação de Mestrado, apresentada na Universidade de Lisboa - FCUL. Lisboa.
- [33] Palma, W. (2007). *Long-Memory Time Series - Theory and Methods*. New Jersey: John Wiley.
- [34] Rousseeuw, P.J. e A.M. Leroy (2003). *Rubust Regression and Outlier Detection*. New Jersey: John Wiley.
- [35] Rudin, W. (1976). *Principles of Mathematical Analysis*. Auckland: Mcgraw-Hill.
- [36] Sowell, F. (1992). "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models". *Journal of Econometrics*, Vol. **53**, pp. 165-188.

- [37] Taqqu, M.S., V. Teverovsky e W. Willinger (1995). “Estimators for Long Range Dependence: An Empirical Study”. *Fractals*, Vol. **3**(4), pp. 785-798.
- [38] Tsay, R.S. (1986). “Time Series Model Specification in the Presence of Outliers”. *Journal of the American Statistical Association*, Vol. **81**, pp. 132-141.
- [39] Wei, W. (2005). *Time Series Analysis. Univariate and Multivariate Methods*. New York: Addison
- [40] W. S. Cleveland, E. Grosse and W. M. Shyu (1992) “Local regression models.” Chapter 8 of *Statistical Models in S* eds J.M. Chambers and T.J. Hastie, Wadsworth and Brooks/Cole.Wesley.
- [41] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. New York: Hafner.
- [42] Yohai, V. J. (1987). “High breakdown point and high efficiency robust estimates for regression”. *Annals of Statistics*, Vol. **15**, pp. 642-656.

Apêndice A

As Tabelas A.1 a A.12 a seguir apresentam os resultados de estimação paramétrica dos processos ARFIMA(p, d, q), quando $d \in \{0.1, 0.3, 0.45\}$, $p \in \{0, 1\}$, $q \in \{0, 1\}$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$ através do modelo de contaminação por mistura com probabilidade de contaminação $p = 0.05$ no caso de *outliers* dos tipos *AO* e *IO*, considerando os estimadores *FT*, *FTAO* e *FTIO*.

Tabela A.1: Estimaco paramtrica dos processos ARFIMA(p, d, q), quando $d = 0.1$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO				FT		
$n = 250$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.1326	0.1696	0.2039	0.0651	0.0537	0.0475
vício	0.0326	0.0696	0.1039	-0.0349	-0.0463	-0.0525
eqm	0.0057	0.0139	0.0229	0.0028	0.0035	0.004
variância	0.0047	0.0091	0.0121	0.0016	0.0014	0.0012
$n = 500$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.1257	0.1508	0.1774	0.0639	0.0464	0.0411
vício	0.0257	0.0508	0.0774	-0.0361	-0.0536	-0.0589
eqm	0.0037	0.0079	0.0142	0.0023	0.0037	0.0042
variância	0.0031	0.0053	0.0082	0.001	0,0008	0,0007
$n = 1000$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.1242	0.1412	0.1608	0.0643	0.0437	0.0342
vício	0.0242	0.0412	0.0608	-0.0357	-0.0563	-0.0658
eqm	0.0023	0.0054	0.0092	0.0019	0.0037	0.0048
variância	0.0017	0.0037	0.0055	0,0006	0,0005	0,0004

Tabela A.2: Estimaco paramtrica dos processos ARFIMA(p, d, q), quando $d = 0.45$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO				FT		
$n = 250$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.4352	0.4192	0.4109	0.3243	0.2332	0.1735
vício	-0.0148	-0.0308	-0.0391	-0.1257	-0.2168	-0.2765
eqm	0.003	0.0055	0.0081	0.0192	0.0504	0.0798
variância	0.0028	0.0045	0.0066	0.0034	0.0034	0.0033
$n = 500$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.4547	0.4514	0.4438	0.328	0.2457	0.1878
vício	0.0047	0.0014	-0.0062	-0.122	-0.2043	-0.2622
eqm	0.0012	0.0017	0.0026	0.0165	0.0435	0.0706
variância	0.0012	0.0017	0.0026	0.0016	0.0018	0.0019
$n = 1000$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.4658	0.4655	0.4621	0.3349	0.249	0.1987
vício	0.0158	0.0155	0.0121	-0.1151	-0.201	-0.2513
eqm	0.0009	0.001	0.0011	0.014	0.0413	0.0641
variância	0.0006	0.0008	0.001	0.0008	0.0009	0.0009

Tabela A.3: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1$, $p = 1$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.1213	-0.6523	0.1742	-0.6911	0.2169	-0.7195	0.0990	-0.4504	0.0869	-0.3347	0.0790	-0.2623	
vício	0.0213	-0.0523	0.0742	-0.0911	0.1169	-0.1195	-0.0010	0.1496	-0.0131	0.2653	-0.0210	0.3377	
eqm	0.0061	0.0029	0.0157	0.0087	0.0287	0.0145	0.0031	0.0276	0.0032	0.0768	0.0031	0.1198	
variância	0.0057	0.0002	0.0102	0.0004	0.0150	0.0003	0.0031	0.0052	0.0030	0.0064	0.0027	0.0058	
$n = 500$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.0970	-0.6488	0.1334	-0.6864	0.1946	-0.7102	0.0953	-0.4501	0.0828	-0.3333	0.0686	-0.2492	
vício	-0.0030	-0.0488	0.0334	-0.0864	0.0946	-0.1102	-0.0047	0.1499	-0.0172	0.2667	-0.0314	0.3508	
eqm	0.0034	0.0024	0.0076	0.0077	0.0223	0.0162	0.0020	0.0252	0.0023	0.0746	0.0028	0.1266	
variância	0.0034	0.0001	0.0065	0.0002	0.0134	0.0040	0.0020	0.0027	0.0020	0.0035	0.0018	0.0035	
$n = 1000$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.0840	-0.6464	0.1052	-0.6798	0.1549	-0.7092	0.0995	-0.4534	0.0842	-0.3315	0.0657	-0.2473	
vício	-0.0160	-0.0464	0.0052	-0.0798	0.0549	-0.1092	-0.0005	0.1466	-0.0158	0.2685	-0.0343	0.3527	
eqm	0.0022	0.0022	0.0041	0.0065	0.0126	0.0140	0.001	0.023	0.0015	0.0740	0.0022	0.1263	
variância	0.0019	0.0000	0.0041	0.0001	0.0096	0.0020	0.0010	0.0015	0.0013	0.0019	0.0011	0.0019	

Tabela A.4: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45$, $p = 1$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.4202	-0.6128	0.3928	-0.6198	0.3681	-0.5506	0.3142	-0.4163	0.2172	-0.2800	0.1647	-0.2073	
vício	-0.0298	-0.0128	-0.0572	-0.0198	-0.0819	0.0494	-0.1358	0.1837	-0.2328	0.3200	-0.2853	0.3927	
eqm	0.0049	0.0005	0.0114	0.0016	0.0193	0.0643	0.0236	0.0403	0.0604	0.1109	0.0872	0.1628	
variância	0,004	0,0004	0.0082	0.0012	0.0126	0.0619	0.0052	0.0066	0.0062	0.0085	0.0057	0.0086	
$n = 500$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.4471	-0.6276	0.4366	-0.6398	0.4116	-0.5984	0.3326	-0.4328	0.2401	-0.3001	0.1800	-0.2165	
vício	-0.0029	-0.0276	-0.0134	-0.0398	-0.0384	0.0016	-0.1174	0.1672	-0.2099	0.2999	-0.2700	0.3835	
eqm	0.0014	0.0009	0.0032	0.0022	0.0085	0.0305	0.0164	0.0311	0.0472	0.0943	0.0758	0.1510	
variância	0.0014	0.0002	0.0031	0.0006	0.0071	0.0305	0.0027	0.0032	0.0031	0.0043	0.0029	0.0040	
$n = 1000$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.4643	-0.6364	0.4610	-0.6548	0.4509	-0.6495	0.3419	-0.4399	0.2571	-0.3142	0.1966	-0.2319	
vício	0.0143	-0.0364	0.0110	-0.0548	0.0009	-0.0495	-0.1081	0.1601	-0.1929	0.2858	-0.2534	0.3681	
eqm	0.0009	0.0014	0.0010	0.0033	0.0018	0.0098	0.0128	0.0271	0.0385	0.0836	0.0658	0.1377	
variância	0,0007	0,0001	0,0009	0,0003	0.0018	0.0074	0.0011	0.0014	0.0013	0.0019	0.0016	0.0022	

Tabela A.5: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1$, $p = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.1130	-0.7355	0.1465	-0.7425	0.1830	-0.7085	0.0744	-0.3314	0.0676	-0.1952	0.0608	-0.1251	
vício	0.0130	-0.1355	0.0465	-0.1425	0.0830	-0.1085	-0.0256	0.2686	-0.0324	0.4048	-0.0392	0.4749	
eqm	0.0036	0.0189	0.0076	0.0222	0.0142	0.0178	0.0026	0.0723	0.0026	0.1640	0.0029	0.2256	
variância	0.0034	0.0005	0.0054	0.0019	0.0073	0.0060	0.0019	0.0001	0.0016	0.0001	0.0014	0.0001	
$n = 500$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.1247	-0.6721	0.1449	-0.7321	0.1620	-0.7561	0.0751	-0.3266	0.0627	-0.1999	0.0543	-0.1300	
vício	0.0247	-0.0721	0.0449	-0.1321	0.0620	-0.1561	-0.0249	0.2734	-0.0373	0.4001	-0.0457	0.4700	
eqm	0.0031	0.0092	0.0061	0.0331	0.0096	0.0630	0.0019	0.0751	0.0026	0.1606	0.0031	0.2213	
variância	0.0025	0.0040	0.0040	0.0157	0.0057	0.0387	0.0013	0.0004	0.0012	0.0005	0,001	0,0004	
$n = 1000$													
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	
w	4		7		10		4		7		10		
média	0.1013	-0.7705	0.1117	-0.8567	0.1609	-0.7555	0.0785	-0.3264	0.0609	-0.1982	0.0520	-0.1336	
vício	0.0013	-0.1705	0.0117	-0.2567	0.0609	-0.1555	-0.0215	0.2736	-0.0391	0.4018	-0.0480	0.4664	
eqm	0.0010	0.0293	0.0016	0.0668	0.0066	0.0383	0.0011	0.0749	0.0022	0.1615	0.0029	0.2176	
variância	0,001	0,0003	0.0015	0.0009	0.0029	0.0142	0,0007	0,0001	0,0006	0,0001	0,0006	0,0001	

Tabela A.6: Estimaco paramtrica dos processos ARFIMA(p, d, q), quando $d = 0.45$, $p = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT					
n = 250												
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$
w	4		7		10		4		7		10	
média	0.4172	-0.7242	0.4150	-0.7692	0.4114	-0.7658	0.3791	-0.2368	0.3345	-0.0826	0.2847	-0.0126
vício	-0.0328	-0.1242	-0.0350	-0.1692	-0.0386	-0.1658	-0.0709	0.3632	-0.1155	0.5174	-0.1653	0.5874
eqm	0.0042	0.0159	0.0048	0.0300	0.0060	0.0303	0.0085	0.1321	0.0173	0.2680	0.0312	0.3453
variância	0.0032	0.0005	0.0036	0.0014	0.0045	0.0028	0.0035	0.0002	0,004	0,0002	0.0039	0.0002
n = 500												
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$
w	4		7		10		4		7		10	
média	0.4352	-0.7084	0.4311	-0.7759	0.4344	-0.7957	0.4062	-0.2087	0.3581	-0.0556	0.3196	0.0199
vício	-0.0148	-0.1084	-0.0189	-0.1759	-0.0156	-0.1957	-0.0438	0.3913	-0.0919	0.5444	-0.1304	0.6199
eqm	0.0021	0.0247	0.0024	0.0324	0.0024	0.0419	0.0038	0.1537	0.0104	0.2966	0.0189	0.3845
variância	0.0019	0.0130	0.0020	0.0015	0.0022	0.0036	0.0019	0.0006	0,002	0,0002	0.0019	0.0002
n = 1000												
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$
w	4		7		10		4		7		10	
média	0.4479	-0.6920	0.4454	-0.7658	0.4466	-0.8049	0.4138	-0.1998	0.3744	-0.0380	0.3352	0.0351
vício	-0.0021	-0.0920	-0.0046	-0.1658	-0.0034	-0.2049	-0.0362	0.4002	-0.0756	0.5620	-0.1148	0.6351
eqm	0.0009	0.0087	0.0011	0.0281	0.0012	0.0435	0.0022	0.1602	0.0067	0.3160	0.0141	0.4035
variância	0,0009	0,0003	0.0011	0.0006	0.0012	0.0016	0,0009	0,0001	0,001	0,0001	0,0009	0,0001

Tabela A.7: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO				FT		
$n = 250$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.1006	0.1006	0.1018	0.0887	0.0885	0.0897
vício	0,0006	0,0006	0.0018	-0.0113	-0.0115	-0.0103
eqm	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002
variância	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
$n = 500$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.0986	0.0997	0.0997	0.0914	0.0994	0.0994
vício	-0.0014	-3E-04	-3E-04	-0.0086	-6E-04	-6E-04
eqm	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000
variância	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
$n = 1000$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.0985	0.0996	0.0999	0.0945	0.0955	0.0959
vício	-0.0015	-4E-04	-1E-04	-0.0055	-0.0045	-0.0041
eqm	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000
variância	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela A.8: Estimaco paramtrica dos processos ARFIMA(p, d, q), quando $d = 0.45$, $p = 0$, $q = 0$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO				FT		
$n = 250$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.44	0.4408	0.4386	0.4309	0.4313	0.43
vício	-0.01	-0.0092	-0.0114	-0.0191	-0.0187	-0.02
eqm	0,0002	0,0002	0,0002	0,0004	0,0004	0,0005
variância	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
$n = 500$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.4471	0.4486	0.4484	0.4469	0.4483	0.4482
vício	-0.0029	-0.0014	-0.0016	-0.0031	-0.0017	-0.0018
eqm	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
variância	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
$n = 1000$						
	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}	\hat{d}
w	4	7	10	4	7	10
média	0.4519	0.451	0.4515	0.4482	0.4474	0.4479
vício	0.0019	0.001	0.0015	-0.0018	-0.0026	-0.0021
eqm	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
variância	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabela A.9: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.1$, $p = 1$, $\phi_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT						
$n = 250$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.0999	-0.5976	0.1011	-0.5982	0.0987	-0.5970	0.0984	-0.5969	0.1013	-0.5976	0.0980	-0.5964	
vício	-0.0001	0.0024	0.0011	0.0018	-0.0013	0.0030	-0.0016	0.0031	0.0013	0.0024	-0.0020	0.0036	
eqm	0.0025	0.0001	0.0025	0.0001	0.0025	0.0001	0.0025	0.0001	0.0024	0.0001	0.0024	0.0001	
variância	0.0025	0.0001	0.0025	0.0001	0.0025	0.0001	0.0025	0.0001	0.0024	0.0001	0.0024	0.0001	
$n = 500$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.0955	-0.5967	0.0972	-0.5980	0.0985	-0.5964	0.0948	-0.5964	0.0966	-0.5977	0.0980	-0.5961	
vício	-0.0045	0.0033	-0.0028	0.0020	-0.0015	0.0036	-0.0052	0.0036	-0.0034	0.0023	-0.0020	0.0039	
eqm	0.0013	0.0001	0.0012	0.0001	0.0013	0.0001	0.0014	0.0001	0.0013	0.0001	0.0013	0.0001	
variância	0.0013	0.0001	0.0012	0.0001	0.0013	0.0001	0.0013	0.0001	0.0012	0.0001	0.0013	0.0001	
$n = 1000$													
	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	\hat{d}	$\hat{\phi}_1$	
w	4		7		10		4		7		10		
média	0.0985	-0.5976	0.0980	-0.5986	0.0978	-0.5968	0.0983	-0.5974	0.0977	-0.5984	0.0975	-0.5967	
vício	-0.0015	0.0024	-0.0020	0.0014	-0.0022	0.0032	-0.0017	0.0026	-0.0023	0.0016	-0.0025	0.0033	
eqm	0,0006	0,0000	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000	0,0006	0,0000	0,0007	0,0000	
variância	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	0,0006	0,0000	0,0007	0,0000	

Tabela A.12: Estimação paramétrica dos processos ARFIMA(p, d, q), quando $d = 0.45$, $p = 1$, $\theta_1 = -0.6$, $n \in \{250, 500, 1000\}$ e $w \in \{4, 7, 10\}$.

FTAO							FT					
$n = 250$												
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$
w	4		7		10		4		7		10	
média	0.4290	-0.6005	0.4291	-0.6015	0.4298	-0.6015	0.4284	-0.6008	0.4286	-0.6020	0.4294	-0.6021
vício	-0.0210	-0.0005	-0.0209	-0.0015	-0.0202	-0.0015	-0.0216	-0.0008	-0.0214	-0.0020	-0.0206	-0.0021
eqm	0.0025	0.0001	0.0025	0.0001	0.0024	0.0001	0.0025	0.0001	0.0026	0.0001	0.0025	0.0001
variância	0.0021	0.0001	0.0021	0.0001	0,002	0,0001	0.0021	0.0001	0.0021	0.0001	0,002	0,0001
$n = 500$												
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$
w	4		7		10		4		7		10	
média	0.4426	-0.5978	0.4424	-0.5972	0.4412	-0.5980	0.4424	-0.5981	0.4418	-0.5977	0.4406	-0.5984
vício	-0.0074	0.0022	-0.0076	0.0028	-0.0088	0.0020	-0.0076	0.0019	-0.0082	0.0023	-0.0094	0.0016
eqm	0.0011	0.0001	0.0012	0.0001	0.0012	0.0001	0.0011	0.0001	0.0012	0.0001	0.0012	0.0001
variância	0.0011	0.0001	0.0011	0.0001	0.0011	0.0001	0.0011	0.0001	0.0011	0.0001	0.0012	0.0001
$n = 1000$												
	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$	\hat{d}	$\hat{\theta}_1$
w	4		7		10		4		7		10	
média	0.4502	-0.5974	0.4504	-0.5972	0.4484	-0.5979	0.4497	-0.5976	0.4499	-0.5973	0.4480	-0.5981
vício	0.0002	0.0026	0.0004	0.0028	-0.0016	0.0021	-0.0003	0.0024	-0.0001	0.0027	-0.0020	0.0019
eqm	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000
variância	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000	0,0006	0,0000	0,0007	0,0000	0,0006	0,0000